Detection and Characterisation of Quantitative Trait Loci Affecting Muscle and Growth Phenotypes in Sheep

Georgia Hadjipavlou

This thesis is presented for the degree of

Doctor of Philosophy

University of Edinburgh

2010

Σαν έξαφνα, ώρα μεσάνυχτ', ακουσθεί
αόρατος θίασος να περνά
με μουσικές εξαίσιες, με φωνές—
την τύχη σου που ενδίδει πια, τα έργα σου
που απέτυχαν, τα σχέδια της ζωής σου
που βγήκαν όλα πλάνες, μη ανωφέλετα θρηνήσεις.
Σαν έτοιμος από καιρό, σα θαρραλέος,
αποχαιρέτα την, την Αλεξάνδρεια που φεύγει.
Προ πάντων να μη γελασθείς, μην πεις πως ήταν
ένα όνειρο, πως απατήθηκεν η ακοή σου·
μάταιες ελπίδες τέτοιες μην καταδεχθείς.
Σαν έτοιμος από καιρό, σα θαρραλέος,
σαν που ταιριάζει σε που αξιώθηκες μια τέτοια πόλι,
πλησίασε σταθερά προς το παράθυρο,
κι άκουσε με συγκίνησιν, αλλ' όχι
με των δειλών τα παρακάλια και παράπονα,
ως τελευταία απόλαυσι τους ήχους,
τα εξαίσια όργανα του μυστικού θιάσου,
κι αποχαιρέτα την, την Αλεξάνδρεια που χάνεις.

"Απολείπειν ο θεός Αντώνιον"

Κωνσταντίνος Π. Καβάφης

When abruptly, at midnight, you hear
an invisible procession pass by
with delightful music, and voices,
don't grieve for your failing fortunes,
your spoiled deeds, the illusion of
your life's plan; to moan is useless.
Rather, with foreknowledge and boldness,
bid farewell to the departing Alexandria.
Above all, don't fool yourself, don't claim
it was just a dream, that you heard a lie;
avoid all such futile notions.
As if long prepared, and ever courageous,
acting as one who deserves such a city,
make your way to the window,
and listen closely with your heart, not
with cowardly pleas and protests;
hear, as a last pleasure, those sounds,
the delightful music of the invisible procession,
and bid farewell to the Alexandria you are losing.

"God Abandons Antony"

Constantine P. Kavafy (Translated by Stratis Haviaras)

This thesis is dedicated to my mother

TABLE OF CONTENTS

DECLARATION

I hereby declare that I am the author of this thesis and that I did all the work described herein, unless otherwise specified. In particular to chapter 4, my own contributions to the research are explicitly stated, and work done by others in the collaborative projects described is clearly noted in the text. This thesis is an account of work conducted by me whilst studying for the degree of Doctor of Philosophy at the University of Edinburgh.

G. Hadjipavlou

Georgia Hadjipavlou

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Steve Bishop, for providing an excellent environment for my PhD studies, and making the entire experience fruitful and enjoyable. Steve's scientific thinking has been inspirational and his teaching, guidance and encouragement have been instrumental in the entire course of my work. Because of Steve, my PhD time has been full of stimulating discussions and boundless opportunities for developing my skills as a researcher. His supervision was effective, facilitating structured yet stress-free collaboration and a friendly atmosphere. My appreciation also goes to my university supervisor, Sara Knott, for useful discussions and helpful advice at all stages of my studies.

This work was conducted as part of the GENACT Project, funded by the Marie Curie Host Fellowships for Early Stage Research Training, within the 6th Framework Programme of the European Commission. Special mention goes to the project coordinator, Andrew Leigh Brown, for assistance at various instances.

My gratitude extends to Christos Papachristoforou and Alkis Koumas at the Agricultural Research Institute in Cyprus for their encouragement to apply for a PhD position at the Roslin Institute and for their support throughout my studies.

I am thankful to Oswald Matika for guidance and support during the early stages of my studies, and for a great friendship. My gratitude also goes to Ricardo Pong-Wong for help at many occasions and a productive collaboration. Additionally, I acknowledge effective collaboration with Gib Hemani, Richard Leach, Bruno Louro, Javad Nadaf, Suzanne Rowe and DJ de Koning. Many thanks to Alex Clop, Ross Houston and Wenhua Wei for valuable contributions to my PhD experience.

Special mention is reserved for fellow students, Andreas, Elina, Cecile, Claudia, Kath, Alex, Jen, Hans and Craig (to name a few) for a lively working environment, memorable times and lasting friendships. My gratitude extends to other friends I met in Edinburgh, and in particular to Dave, Daphne, Demos, Elena, Chrysa, Giannis, Dionysia and Katerina.

I would also like to thank dear friends from home for constant support and wonderful times. My “sisters”, Natalie and Marianthi have been an integral part of my life, never missing a beat. Anthi and Alkystis have been beside me with valuable input and humour. Many more friends and relatives are always on my mind. Special mentions go to Lambros, Stella, Giannis, Giota, and to Eleni, Stavros, Katerina for support and care.

I am forever indebted to my mother, Eleftheria Siepi, for her unending love, faith and support. Without her I would not be who I am. I am grateful to my father, Stelios Hadjipavlou, for his unconditional love and encouragement. My wonderful brothers have enriched life every step of the way. Lambros has been there from the early years, providing brotherly fun and devotion. Georgios, Stefanos and Anastasis have brought even more diversity and smiling faces in the family. Panagiotis has been inspiring and a wonderful distraction all this time.

Last but not least, I would like to express my gratitude to Manolis for encouragement, confidence and endless optimism. Without his love and energy, life during the PhD would not have been as fulfilling and meaningful.

LIST OF PUBLICATIONS

Peer-reviewed:

Hadjipavlou G., Matika O., Clop A. & Bishop S.C. 2008. Two single nucleotide polymorphisms in the *myostatin (GDF8)* gene have significant association with muscle depth of commercial Charollais sheep. *Animal Genetics* **39**, 346-353. (Based on **Chapter 2**).

Hadjipavlou G. & Bishop S.C. 2009. Age-dependent quantitative trait loci affecting growth traits in Scottish Blackface sheep. *Animal Genetics* **40**, 165-175. (Based on **Chapter 3**).

Hadjipavlou G., Hemani G., Leach R., Louro B., Nadaf J., Rowe S. & de Koning D.J. Extensive QTL and association analyses of the QTLMAS2009 data. *BMC Proceedings* **4**(Suppl. 1), S11. (Includes the first analysis described in **Chapter 4**).

Pong-Wong R. & **Hadjipavlou G**. A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data. *BMC Proceedings* **4**(Suppl. 1), S4. (Based on the second analysis described in **Chapter 4**).

Conference Proceedings:

Hadjipavlou G., Matika O., Clop A. & Bishop S.C. Two myostatin single nucleotide polymorphisms have significant effects on carcass traits of UK commercial Charollais sheep populations. *Proceedings of the 3rd International Conference in Quantitative Genetics, Zhejiang University, Hangzhou, China*, 19-24 August 2007.

Hadjipavlou G., Matika O., Clop A. & Bishop S.C. 2008. Two *myostatin* single nucleotide polymorphisms have significant effects on muscle depth of British commercial Charollais sheep. *Proceedings of the British Society of Animal Science* Abstract No. 68.

Hadjipavlou G. & Bishop S.C. 2008. Dissection of quantitative trait loci associated with growth-related traits in Scottish Blackface sheep. *Proceedings of the 59th Annual Meeting of the European Association of Animal Production (EAAP)*, *Vilnius, Lithuania*, 24-27 August 2008, Abstract No.3046.

Hadjipavlou G. & Bishop S.C. 2009. Application of random regression techniques to dissect age-dependent quantitative trait loci for growth in lambs. *Proceedings of the Association for Advancement of Animal Breeding and Genetics (AAABG)* **18**, 446-449. (Based on **Chapters 5** and **6**).

ABSTRACT

This thesis addresses the dissection and characterisation of quantitative trait loci (QTL) affecting production traits in sheep.

Firstly, the association between specific genetic polymorphisms and complex variation in weight, muscle and fat depositions was investigated. Research concentrated on assessing the presence, correspondence and significance of two single nucleotide polymorphisms (SNPs) in the *GDF8* region of ovine chromosome 2, reportedly affecting muscle production. Commercial populations of British Texel, Suffolk and Charollais sheep were studied. The SNPs were absent in Suffolk and almost fixed in Texel breeds. In the Charollais population, the SNPs segregated at intermediate frequencies and a significant association was found between these polymorphisms and muscle depth. The previously proposed causative allele at one of the loci resulted in increased muscle depth and, at allele frequency of 0.5, this locus would explain one third of the additive genetic variance for the trait. Partial recessive allelic expression is proposed by genotypic value predictions and is consistent with the previously postulated molecular mechanism by which it gives rise to muscle changes.

Secondly, the thesis focused on detection of QTL associated with growth. Live weight is a composite of growth rates over time, with inter-age genetic correlations for live weight decreasing as time between weight measurements increases. To explore whether observed genetic correlation patterns translate into distinct loci acting on weight at different growth stages, a novel method was

developed and the applicability of a second proposed method was explored. Both methods allowed simultaneous analysis of multiple live weights per animal, while accounting differently for the correlation among measurements ordered in time.

In the first approach, a growth curve technique was developed and employed to map growth QTL for curve parameters and predicted growth descriptors. A study of actual live weights identified significant QTL at different ages on distinct chromosomes, with QTL significance and variance changing over time. Further application of this technique on a simulated dataset validated its effectiveness in detecting age-dependent QTL. An extension of the procedure resulted in a novel technique for genomic evaluation of longitudinal traits.

In the second method examined, random regression (RR) models were applied for dissection of growth QTL. Systematic model selection and inclusion of relevant random effects resulted in apparently significant QTL, but the method was computationally demanding, model choice proved challenging and the results were questioned. To further explore the method, RR models were applied to various simulated growth phenotypes composed of time-dependent QTL trajectories, polygenic and environmental effects. Statistically optimal RR models succeeded in identifying significant QTL and predicting the simulated time-dependence for most scenarios. However, the issue of model choice was again prominent, as suboptimal models resulted in unreliable QTL variance trajectories and pronounced confounding between different time-dependent effects. Thus, the growth curve approach appeared to be the more flexible and robust process for analysing longitudinal data to map age-dependent QTL.

CHAPTER 1

General Introduction

1.1 Overview

In the past 20 years, a combination of advances in molecular genetics and improvements in statistical and computational techniques occurred. This led to a vast number of studies aiming at dissecting quantitative traits of biological, biomedical or agricultural importance into genetic loci, termed quantitative trait loci (QTL), that explain a proportion of the phenotypic variance for the trait.

In livestock, QTL mapping experiments have identified chromosomal regions that contain loci of potential commercial benefit. QTL have been mainly identified by utilising the principle of co-segregation between putative QTL and specific molecular marker alleles due to their physical linkage. However, full characterisation of the QTL effects on economically important traits has often not been feasible. The primary reason for this is that, in the majority of cases, QTL were detected in experimental crosses of divergent lines (backcross, F2, full-sib, and half-sib families), with co-segregation between specific QTL and marker alleles not replicated in other lines or commercial populations of the same breed. Because of this, further QTL experiments have had to be performed in candidate chromosomal regions in commercial populations of interest (Nagamine *et al.* 2003; de Koning *et al.* 2003; Walling *et al. 2004)*, with the ultimate objective of fine-mapping the QTL and dissecting the underlying gene(s) or polymorphism(s) that affected the trait under

study. However, identification and verification of the causative variants for the QTL has been successful in a small number of cases (e.g. Wilson *et al.*, 2001; Grisart *et al.*, 2002; Van Laere *et al.,* 2003; Clop *et al.*, 2006). When experiments succeed in fine-mapping a QTL or identifying the genetic loci causing phenotypic change, validation, description and quantification of QTL effects in commercial breeds, although challenging, is of primary interest. Characterisation of the effects of a causative locus on a trait in the relevant population would allow full evaluation of its importance and benefits for the livestock industry, and possibly enable direct incorporation of the findings in a breed's genetic selection programme.

An additional challenge for QTL mapping experiments is the dissection of loci that influence longitudinal traits, i.e. the genetic and environmental components of which change with time. Growth and its manifestation, live weight, are such longitudinal traits of major importance for livestock species. Although an easily recorded phenotype, live weight is a composite trait as it is the integral of all growth rates prior to that point, which in turn are a function of the animal's health, physiological state, strength of immune system and even its ability to compete for sometimes limited nutritional resources. Therefore, it would be beneficial to determine aspects of growth governed by genetic effects, and to utilize this information for genetic improvement.

Univariate QTL analyses of sheep live weight phenotypes at a particular point in time have been performed previously (e.g. Walling *et al.* 2004; McRae *et al.* 2005). However, isolation of live weights as single traits fails to capture the correlations among the components underlying growth. As a result, univariate studies have reduced power to detect QTL compared to techniques that combine information

from multiple phenotypes (Jiang & Zeng 1995; Knott & Haley 2000; Sorensen *et al.* 2003). Some multivariate approaches for identifying QTL have involved the transformation of multiple traits into single summary or composite measures (principal component analysis). With this method, Weller *et al.* (1996) analysed canonical variables to map QTL for milk production traits in cattle and Stearns *et al.* (2005) found QTL for carcass and meat quality traits in swine. In other studies, single-trait maximum likelihood or regression methods for QTL mapping were simply extended to multiple traits (Jiang & Zeng 1995; Wu *et al.* 1999; Knott & Haley 2000). In these studies no structure (such as time dependence) was assumed among the traits analysed. Generally, an increase in QTL detection power was observed in all the above methods because information from multiple traits was used simultaneously. However, their applications have been restricted to bivariate or trivariate analyses, with several concerns being raised when more traits are included. Decreased precision in genetic parameter estimation and complex multi-trait test statistics for testing QTL significance and location are among the unresolved issues (Ma *et al.* 2002; Knott 2005). Thus, the practical usefulness of full multivariate methods for longitudinal trait analysis is limited.

A few studies have tried to account for time-dependent QTL effects on phenotypes recorded at several time points. Methods that used non-linear functions to model QTL as fixed effects in mixed models have been proposed (Ma *et al.* 2002; Wu & Hou 2006) and applied to map QTL for traits in various species (Rodriguez-Zas *et al.* 2002; Wu *et al.* 2004a,b; Wu *et al.* 2005; Minvielle *et al.* 2006). In most of these studies, polygenic and environmental components were estimated as aggregates in the residual and were assumed to be time independent. Alternatively, the QTL was

fitted as a random effect using random regression in a small number of simulation studies (Lund *et al.* 2002; Macgregor *et al.* 2005) and a single project that used actual longitudinal data (Lund et a*l.* 2008). The latter study modelled time-dependent polygenic and environmental effects. Overall, all the above approaches provided possible routes for QTL detection in longitudinal data. Yet a number of issues remain prior to further applying these techniques, such as optimal model choice, adequate description of time-dependent QTL and other effects, computational time and statistical difficulties in estimating many parameters simultaneously.

This thesis focuses on dissecting genetic aspects of traits that influence sheep muscle composition and growth, these being the major traits (along with disease resistance) affecting lamb performance. Firstly, further exploration and description of recent findings from QTL fine-mapping experiments regarding causative variants affecting muscularity is pursued. Secondly, great emphasis is placed on the development and application of two different techniques for efficient analysis of longitudinal growth phenotypes in order to detect and characterise growth QTL. Below, the thesis rationale is explained through an account of pertinent literature findings on the genetic basis of muscularity and growth in sheep. The specific research objectives are described in more detail in the final section of this chapter.

1.2 Genetic studies on muscle composition in sheep

Intensive studies of major genes connected with sheep muscle and fat composition have focused on the region of ovine chromosome 18 (OAR18) containing the *callipyge* (Cockett *et al.* 1994) and rib-eye muscling (*REM* or *Carwell*) loci (Nicoll *et al.* 1998), or the region of OAR2 containing the growth

differentiation factor 8 (*GDF8*), also known as the myostatin (*MSTN)* gene, which is responsible for double muscling in cattle breeds (McPherron & Lee 1997; Kambadur *et al.* 1997; Wiener *et al.* 2002). In sheep, QTL studies showed that a portion of the OAR2 that included *GDF8* had a major effect on muscular development, reminiscent of the cattle double-muscling phenotype, in Belgian Texel (Marcq *et al.* 2002), on muscling and fat depth in New Zealand Texel sires (Broad *et al.* 2000; Johnson *et al.* 2005), British Texel (Walling *et al.* 2004) and Charollais (McRae *et al.* 2005) sheep. Interestingly, no sequence differences were found between the *GDF8* coding sequence of double-muscled Belgian Texels and normally muscled Romanov controls (Marcq *et al.* 2002). This indicated that the functional polymorphism resided outside the *GDF8* coding segment or in a closely linked gene.

Recently, the genetic basis of *GDF8* effects on muscle growth in Texel sheep appears to have been elucidated. Examination of a 10.5 kb gDNA region spanning *GDF8* led to the identification of two biallelic single nucleotide polymorphisms (SNPs) with significantly different allelic frequencies between hyper-muscled Texel and control animals (Clop *et al.* 2006). The first SNP (*g-2449G>C*) was located 2.5 kb upstream from the *GDF8* transcription start site. This SNP and its effect on *GDF8* function (and hence on muscle and fat growth) was not studied any further. The second SNP (*g+6723G>A*) was found in the 3'-UTR of *GDF8*. Functional studies provided evidence that the presence of the A allele at this nucleotide position creates a miRNA target site. This, in turn, leads to miRNA-mediated translational inhibition of *GDF8* by which the double muscling phenotype arises (Clop *et al.* 2006). Thus, the *GDF8 g+6723A* allele seemed to act as a causative variant of increased

muscularity in Belgian Texel rams and could be identified as a quantitative trait nucleotide (QTN).

In addition to the *g+6723A* QTN, a nucleotide deletion (*c.960delG*) and insertion (*c.120insA*) at two distinct loci in the *GDF8* coding region were lately proposed as causative for carcass conformation and fatness in Norwegian White (Boman *et al.* 2009) and Norwegian Spælsau sheep (Boman & Våge 2009), respectively. Examination of haplotypes including different allelic combinations at *c.960delG* and *g+6723G>A* indicated that the deletion mutation had more profound effects on phenotypes than the 3'-UTR QTN, first reported by Clop *et al.* (2006).

Full characterization of a presumed QTN or causative variant requires assessing its presence, correspondence and significance for traits of interest across commercial populations and breeds. Additionally, validation of a causative variant in different breeds is necessary to determine the consequences of different genetic backgrounds and environmental influences on causative variant modulation. Indeed, the influence of background genetics on the effects of *GDF8* mutations on phenotypes has been shown in cattle (Short *et al.* 2002) and mice (Bünger *et al.* 2004; Rehfeldt *et al.* 2005). In addition, the effect of overexpression of other molecules interacting with myostatin and other growth factors on muscle development has been shown in mice (Lee 2007).

Genetic background differences among breeds or populations across countries become particularly relevant when a putative causative polymorphism directing change in a quantitative trait may be translated into commercial technology for marker assisted selection. Since marker assisted selection is often implemented across countries, it is essential for the overall causative variant effects to be estimated

in commercial breeds for a specific region, prior to breeders taking up the technology. Only after a variant's relevance and benefits are thoroughly evaluated, can decisions be made regarding QTN incorporation in breeding programmes of a particular country.

Apart from being important for productive animal breeding, findings from extensive across-breed studies of QTN effects on traits could provide further evidence for the validity of a putative QTN and the means by which the proposed biological function of the causative variant gives rise to phenotypic changes. Additionally, the mode of action by which a causative allele alters the trait of interest is of value to fully assess the strategy and benefits from marker assisted selection.

1.3 Growth QTL studies

Several studies in livestock, although few in sheep, have reported QTL associated with growth traits in terms of average daily gain, weight at a specific age, and days to reach a particular weight (e.g. Stone *et al.* 1999 for cattle; Nagamine *et al.* 2003 and Stearns *et al.* 2005 for pigs). The majority of QTL mapping studies used univariate approaches to detect QTL, treating weights recorded at a particular growth point as separate traits. This is despite the fact that live weights across time comprise a longitudinal trait that is a function of several physiological processes and a composite of growth rate phenotypes over time.

Additionally, strong genetic correlations exist among live weights at different ages (Corva & Medrano 2001; Riggio *et al.* 2008), although patterns of correlations often suggest additional complexity. For example, using sheep data, Fischer *et al.* (2004; 2006) and Riggio *et al.* (2008) independently showed that inter-age genetic

correlations for live weight, whilst strongly positive, are often different from unity, with the correlation decreasing as the time between the weight measurements increases. Thus, it is likely that distinct loci act on live weights at different growth stages. For the detection of QTL that are associated with growth, it would be beneficial to simultaneously analyze multiple measurements and take account of the correlation structure of measurements across time.

Fitting growth models on body weight data from different time-points and extracting the relevant growth parameters provides a means to combine phenotypic information from multiple measurements into a few growth variables in a biologically meaningful manner. Flexible sigmoid curves, such as the logistic, the Brody, Bertalanffy, Gompertz and Richards curves, are extensively described in the literature for livestock (e.g. Pittroff *et al.* 2008). Many growth curve variables relevant to genetic studies may be derived from such models, describing growth rates and live weights over time, as well as features such as maximum growth rate, the age at which it is predicted to occur, and mature weight. Growth model fitting and estimation of growth curve parameters have been previously applied (Lopez *et al.* 2000, several species; Schinckel *et al.* 2004, pigs; Lewis *et al.* 2002 and Lambe *et al.* 2006, sheep).

In sheep, there is little published information on the genetic control of growth curve variables and growth, in general, as a trait. A few studies have investigated the polygenic components of growth curve parameters (Lewis *et al.* 2002; Lambe *et al.* 2006). These indicated that growth variables are heritable and that genetic differences exist among growth parameters of various breeds. Moreover, results from Lambe *et al.* (2006) suggested that early growth rate is a different genetic trait to late

growth rate. All these postulations support a hypothesis that growth curve variables are under genetic regulation and that they may constitute separate aspects of the complex, longitudinal trait of growth. Validation of this assertion requires a more detailed description of the genetic control of growth. To this end, it would be informative to dissect the genetic loci that underlie growth curve predictors.

An extension of the growth model applications can be made by treating growth curve variables or predictions as traits for QTL studies. In an analogous manner, non-linear curve parameter estimation was employed by Rodriguez-Zas *et al.* (2002) and Minvielle *et al.* (2006) to detect QTL for dairy cattle milk traits and quail egg production traits, respectively. For some of the QTL, Rodriguez-Zas *et al.* (2002) utilised the lactation equation to estimate trait values (e.g. milk yield, protein percentage) over time for alternative QTL allelic effects. Minvielle *et al.* (2006) identified significant QTL for different parameters of the egg laying curve. In addition to these studies, a Bayesian procedure for QTL detection using prior information obtained from a growth model was applied in pigs (Varona *et al.* 2005). However, none of the studies examined the trends of QTL significance across time, and thus, did not determine whether the mapped QTL had time-dependent effects on the longitudinal traits studied.

Random regression (RR) models were previously applied on longitudinal weights in different species to estimate time-dependent polygenic and environmental effects (e.g. Meyer 1998; Meyer 1999; Fischer *et al.* 2004b; Lambe *et al*. 2006), along with maternal genetic components (e.g. Lewis & Brotherstone 2002; Fischer *et al.* 2006; Molina *et al.* 2007). Yet, only a few studies have proposed RR models for QTL mapping (Lund *et al.* 2002; Macgregor *et al.* 2005; Lund *et al.* 2008). Lund *et*

al. (2002) described RR approaches for animals and Macgregor *et al.* (2005) for human populations. Both conducted simulation studies to assess QTL detection power of RR models fitted to longitudinal data with various time-dependence scenarios for the QTL. In both studies, the QTL was modelled as a random effect and covariance functions were used to describe the relationships between QTL effects at different ages. In addition to the QTL, Lund *et al.* (2002) simulated the same time-dependent polygenic effect for all scenarios, and this was then fitted using RR polynomials in the models tested. On the contrary, no polygenic effects were included and constant environmental effects were assumed over time in Macgregor *et al.* (2005). Recently, Lund *et al.* (2008) extended the RR model of the earlier simulation study, to allow for a genome scan and time-dependent QTL mapping for an actual trait. Their main objective was to test the RR methodology against univariate QTL analysis of milk yield in dairy cattle. In this work, a time-dependent polygenic effect was included in the RR model.

Overall, these simulation studies showed that the RR approach allowed flexible model fitting for longitudinal traits, and resulted in a reduced number of model parameters estimated than a full multivariate analysis. However, it must be borne in mind that in both simulations, a rather simplified biological model was used. Despite the caveats, Macgregor *et al.* (2005) and Lund *et al.* (2008) showed that, when there was substantial QTL variance increase over time, even a first degree RR fit for the QTL led to increased QTL detection power. Thus, RR methodology appears promising for identifying time-dependent QTL when compared to other approaches Also, modelling the QTL effect using a RR polynomial on age could potentially allow prediction of the QTL variance pattern over time (Lund *et al.*

2008). However, models with high order RR polynomials fitted for the QTL effect presented convergence problems, potentially due to data limitations or shortcomings arising from intrinsic properties of the RR process. Moreover, as described above, these studies did not address the issue that effective modelling of other time-dependent effects is necessary for accurate description of QTL effects over time. Further development and evaluation of the RR technique would be needed to examine its potential for identification and description of age-dependent QTL for growth and other longitudinal traits.

1.4 Thesis objectives

This thesis addresses the issues described above and which relate to the analysis and interpretation of QTL and QTN for lamb performance. Specifically, the thesis aims are the following:

The first objective is to determine whether the same genetic mutation(s) underlie OAR2 QTL for muscle and fat traits in British commercial breeds. For this purpose, a genetic study of two *GDF8* SNPs on OAR2, previously identified in Belgian Texel rams (Clop *et al.* 2006), is pursued. Association analysis of the SNP effects on phenotypes in extended commercial populations would allow further evaluation and characterisation of their presumed contribution to muscle and fat phenotypes, with the findings being directly relevant to the sheep industry.

The second thesis objective is to gain better insight into the genetic basis of growth as a process in sheep, by identifying and describing QTL that are linked with longitudinal live weights. For this purpose, a novel method is developed and evaluated, and the applicability of a second proposed method is explored. The first

method employs descriptors of growth derived from fitted growth curves for QTL mapping and characterisation. The second approach aims at further exploring and assessing the application of RR models for growth QTL dissection.

The next paragraphs provide an outline of the contents and study objectives for each of the subsequent chapters.

Chapter 2 describes a study in which the presence, correspondence and significance of two *GDF8* SNPs on OAR2 for weight, muscle and fat depositions is assessed in British commercial Texel, Suffolk and Charollais sheep. Findings on SNP associations with muscle depth in the Charollais breed are described further, and the contribution of the SNP loci to the additive genetic variance for the trait is estimated. Additionally, the mode of allelic expression for one of the SNPs, previously proposed to be a QTN for muscularity, is discussed.

Chapter 3 introduces a method for longitudinal live weight analysis for the purpose of mapping QTL for growth. The process for growth model fitting to longitudinal live weights for Scottish Blackface lambs is presented, along with estimation of individual model parameters and predictions for live weight and growth rate over time. All growth descriptors are employed for QTL mapping, and the age-dependence of identified QTL effects is scrutinized.

Chapter 4 outlines a further application of the technique for growth QTL mapping, described in chapter 3, to an independently simulated dataset. The method's effectiveness in detecting age-dependent QTL is explored. Further, an extension of the procedure resulted in a novel approach for genomic evaluation of longitudinal traits, and this concept is presented.

Chapter 5 is devoted to the application of RR methodology on the same longitudinal live weights employed in chapter 3, for the purpose of identifying age-dependent QTL for growth. A systematic procedure for model selection is described, and the outcomes from employing the full models for QTL mapping on four chromosomes are presented and compared to the findings of the growth model approach, employed in chapter 3.

Chapter 6 details further exploration of the RR procedure for growth QTL mapping, first introduced in chapter 5, using simulated growth phenotypes composed of time-dependent QTL, polygenic and environmental effects. The ability of RR analysis to detect the QTL is assessed for different scenarios of QTL time-dependence. Additionally, the performance of statistically optimal and suboptimal RR models to describe the simulated changes in variance along time for the QTL and other effects is evaluated.

Chapter 7 contains an overall summary and conclusions from this thesis. Fulfilment of the objectives set at the beginning of the work is considered and evaluated, and an overall perspective on the significant contributions made to the field is given, along with possibilities for future research aiming at method advancement and application for dissecting and characterising genetic loci that affect longitudinal traits.

CHAPTER 2

Two single nucleotide polymorphisms in the *myostatin* (*GDF8*) gene have significant association with muscle depth of commercial Charollais sheep

2.1 Introduction

Fat and lean meat depositions are both selection-responsive traits in sheep. However, routine measurement of carcass composition characteristics (via ultrasonic scanning or computer tomography) remains expensive and difficult. Thus, there is a need to dissect the genetic basis of selection-responsive carcass traits in sheep. This can be facilitated by the detection and characterization of major genes, and quantitative trait nucleotides (QTN) affecting muscle and fat traits in sheep.

In the last decade, a large number of experiments have identified chromosomal regions that contain quantitative trait loci (QTL) of commercial benefit in livestock populations. However, identification and verification of the causative variants for the QTL has been successful only in a small number of cases (e.g. Wilson *et al.* 2001; Grisart *et al.* 2002; Van Laere *et al.* 2003; Clop *et al.* 2006). It is particularly important to directly assess the presence, correspondence and significance of identified QTL or QTN across other commercial populations and breeds prior to their incorporation in general breeding programmes for a specific livestock species. Apart from being important for animal breeding, findings from

extensive across-breed studies of QTN effects on traits could provide further evidence for the validity of a putative QTN and the means by which the proposed biological function of the causative variant gives rise to phenotypic changes on the traits of interest.

In sheep, intensive studies of major genes connected with sheep muscle and fat composition have focused on the region of ovine chromosome 18 (OAR18) containing the *callipyge* (Cockett *et al.* 1994) and rib-eye muscling (*REM* or *Carwell*) loci (Nicoll *et al.* 1998), or the region of OAR2 containing the growth differentiation factor 8 (*GDF8*), also known as the myostatin (*MSTN*) gene, which is responsible for double muscling in cattle breeds (McPherron & Lee 1997; Kambadur *et al.* 1997; Wiener *et al.* 2002). In sheep, QTL studies have shown that a portion of the OAR2 that included *GDF8* had a major effect on muscular development in Belgian Texel (Marcq *et al.* 2002), and on muscling and fat depth in New Zealand Texel sires (Broad *et al.* 2000; Johnson *et al.* 2005) and UK Texel (Walling *et al.* 2004) and Charollais (McRae *et al.* 2005) sheep. Yet, no sequence differences were found between the *GDF8* coding sequence of double-muscled Belgian Texels and normally muscled Romanov controls (Marcq *et al.* 2002). This indicated that the functional polymorphism resided outside the *GDF8* coding segment or in a closely linked gene.

Recently, the genetic basis of *GDF8* effects on muscle growth in the Texel sheep appears to have been elucidated. Examination of a 10.5 kb gDNA region spanning *GDF8* (GenBank accession number DQ530260) led to the identification of two biallelic SNPs with significantly different allelic frequencies between hyper-

muscled Texel and control animals (Clop *et al.* 2006). The first SNP (*g-2449G>C*) was located 2.5 kb upstream from the *GDF8* transcription start site. This SNP and its effect on *GDF8* function (and hence on muscle and fat growth) was not studied any further. The second SNP (*g+6723G>A*) was found in the 3'-UTR of *GDF8*. Functional studies provided evidence that the presence of the A allele at this nucleotide position creates a miRNA target site. This, in turn, leads to miRNA-mediated translational inhibition of *GDF8* by which the double muscling phenotype arises (Clop *et al.* 2006). Thus, the *GDF8 g+6723A* allele seems to act as a causative variant of increased muscularity in Texel rams and could be identified as a QTN.

Lately, a nucleotide deletion (*c.960delG*) and insertion (*c.120insA*) at two distinct loci in the *GDF8* coding region were proposed as causative for carcass conformation and fatness in Norwegian White (Boman *et al.* 2009) and Spælsau sheep (Boman & Våge 2009), respectively. Additionally, the Norwegian White breed also segregated for the *g+6723A* QTN (Boman *et al.* 2009), which appeared to have smaller but complementary effects on phenotype.

The first objective of this study was to determine whether the two *GDF8* SNPs reported by Clop *et al.* (2006), and described above, were present in the British commercial Texel, Suffolk and Charollais breeds. Subsequently, association analyses of the SNP effects on phenotypes were performed in an extended British commercial population of Charollais sheep in order to further evaluate the SNP contribution to muscle and fat tissue composition, and to quantify and characterise the SNP effects on these traits in this breed.

2.2 Materials and Methods

2.2.1 Genotyping information

Nine Suffolk, 38 Texel, and 34 Charollais rams sampled across several British commercial flocks were genotyped for the *g+6723G>A* and *g-2449G>C* SNPs observed in the *GDF8* region of OAR2 (GenBank accession number DQ530260) in progeny of Belgian Texel rams displaying muscle hypertrophy (Clop *et al.* 2006). DNA primers used are described in Clop *et al.* (2006). Because some Charollais sires were segregating for the two SNPs (see Results), SNP genotypic data were subsequently obtained on 338 Charollais lambs from 17 paternal half-sib families dispersed in 12 commercial flocks.

2.2.2 Animals and trait information

Standard records (such as parentage, day of birth, sex, flock, litter size born) and phenotypic data on muscle and fat depth ultrasonically scanned at the third lumbar vertebra and on live weight at scanning were provided by Signet (part of technical division of British Meat and Livestock Commission) for 56499 lambs from British commercial Charollais populations that included the 338 genotyped lambs. The 56499 lambs were born from 1990 to 2006, and the 338 genotyped animals were born from 2002 to 2006. The 56499 animals were scanned at a mean age of 22.0 weeks (s.d.=3.2, range=6.3-48.0 weeks). The 338 genotyped animals had a mean age at scanning of 21.4 weeks (s.d.=1.9, range=13.7-27.3 weeks). Complete pedigree information was available for all animals with phenotypic records.

2.2.3 Treatment of data

The distribution of the fat depth measurements was skewed, and, therefore, the data were transformed using a square-root transformation prior to analysis. The live weight at scanning and muscle depth measurements were analysed without any transformation applied since they were normally distributed. Multiple regression and variance component analyses were performed for each of the traits in order to determine significant fixed effects. All fixed-effect models were fitted using the software package R (2006). Significant fixed effects for all traits were sex, litter size at birth, litter size reared, year, flock, and age of dam. The age at scanning (in days) was fitted as a covariate for each trait.

2.2.4 Haplotype reconstruction

The SNP haplotypes for 262 of the 338 genotyped animals were unambiguously reconstructed using R-tools (Pong-Wong *et al.* 2001). This software utilizes marker genotypic data and pedigree information to determine the gametic haplotypes for each animal. The reconstructed haplotypes were used to estimate the linkage disequilibrium (LD) between the alleles at the two *GDF8* SNPs and to perform mixed model association studies (see below) to assess potential effects of the parental origin of SNP alleles on muscle and fat traits.

2.2.5 SNP linkage disequilibrium

Fisher's exact test was used to determine whether the frequencies of the observed haplotypes for the two SNP loci denoted significant LD, i.e. non-random association between alleles at the two SNPs. Haplotype information for the SNP pair was used to determine the extent of linkage disequilibrium (LD) by estimating the correlation, r, between alleles at the two SNPs and its square, r^2:

$$r_{ij} = D_{ij} / [p_i(1-p_i)p_j(1-p_j)]^{1/2} \text{ (Hill \& Robertson 1968)},$$

where $D_{ij}=p_{ij}-p_ip_j$ is the covariance of gametic frequencies, and p_i, p_j, p_{ij} are the frequencies of allele *i* at the first SNP locus, allele *j* at the second SNP locus and haplotype *ij*, respectively .

2.2.6 Mixed model association analysis

The significant fixed effects and covariates were included in mixed model association analysis to determine the effects of each of the SNPs on all traits (live weight at scanning, muscle depth, and fat depth). In addition to the effects described above, the *g+6723G>A* SNP (3 classes; 1=AA, 2=AG, 3=GG)) and the g-2449G>C SNP (3 classes; 1=CC, 2=CG, 3=GG) genotypes were fitted as fixed effects, separately and simultaneously. Two types of association analyses were performed. In the first analysis, a sire model was fitted using the statistical package R (2006) to analyse the trait data from the 338 genotyped animals. In the second analysis, an animal model was fitted using ASREML (Gilmour *et al.* 2002) to analyse all available phenotypic data and pedigree information for 56499 animals. Detailed

information on animals used and data analysed in the two separate mixed models is given in Table 2.1.

In addition to sire or animal identity (polygenic effects), other random effects were sequentially fitted in each model to determine the model best fitting the trait data, and the nested models were compared using the likelihood ratio test. For the sire model analysis, a model in which flock and dam were also fitted as random effects was chosen. For the animal model analysis, the model in which animal (polygenic effects), dam (maternal genetic), and litter (common environment effects) were fitted as random effects was always the model best fitting the data. A more complex model, in which rearing dam (permanent environment effects) was also fitted as a random effect, resulted in a near-zero variance for rearing dam.

Table 2.1 Summary of animals and records used in association analyses of *GDF8 g+6723G>A* and *g-2449G>C* SNP effects on carcass traits of commercial Charollais sheep

	Numbers of:						
Model[1]	Flocks	Sires	Dams	Scan weight records	Muscle depth records	Fat depth records	Genotyping records
Sire	12	17	226	338	338	338	338
Animal	178	2019	20100	56499	56499	56496	338

[1]Sire model refers to the flock/sire/dam model fitted; Animal model refers to the animal/dam/litter model fitted

A fourth class for the ungenotyped animals was included in the animal model analyses of each SNP genotype effect. Fitting the SNP fixed effect after including a fixed effect with two classes, 'genotyped or not', enabled us to utilise all pedigree and phenotypic information to estimate fixed effects and variance components, whilst assessing the significance of the three SNP genotypes at each locus.

Direct and maternal genetic heritabilities for muscle and fat depth were estimated from the variance components arising from these mixed model analyses. In addition, the proportion of phenotypic variance due to common environment (litter effect) was estimated.

2.2.7 Predictions and SNP genotype effects

Predicted trait values for each genotype class of either SNP, including (co)variances for the predictions, and standard error of differences (SED) for contrasts were obtained from the ASREML analyses. The predicted trait values were used to estimate additive and dominance effects on traits for each SNP, and the proportion of additive genetic variance (V_A) for each trait accounted for by the SNPs. The equations used were: Additive effect, a= (AA-GG)/2; Dominance effect, d=AG-[(AA+GG)/2]; % V_A due to SNP = $100[2pq(a+d(q-p))^2]/V_A$, where AA, GG, AG are the predicted trait values for each genotype class, p, q are the allele frequencies at the SNP locus, and V_A is the additive genetic variance of the trait obtained from an animal model analysis ignoring the SNP effects. Standard errors of the additive and dominance effects were constructed from the variance-covariance matrix of the predicted genotype classes, as were the SED for pairwise contrasts of the SNP genotype classes.

2.2.8 Parent of origin effects

Mixed model association analysis of the parent of origin allelic information for each SNP was performed to investigate whether the mode of inheritance of the allele (paternal or maternal origin) had an effect on the traits. Following haplotype reconstruction, the SNP genotype classes were expanded by subdividing the heterozygotes into two classes, according to the parental origin of the *g+6723A* or the *g-2449C* allele. SNP genotype was then fitted as a fixed effect, first in the sire model and then in the animal model. Since trait data from only the 338 genotyped animals were analysed using the sire model, an overall SNP effect with a total of four genotype classes was included in the sire model association analysis. A fifth SNP genotype class corresponding to the ungenotyped animals was included in the animal model analysis, in which phenotypes from 56499 animals (including the 338 genotyped ones) were analysed. The parent of origin effect was assessed by comparing the predicted trait values of the two classes of heterozygotes for each SNP. Hypothesis testing using a t-test was employed.

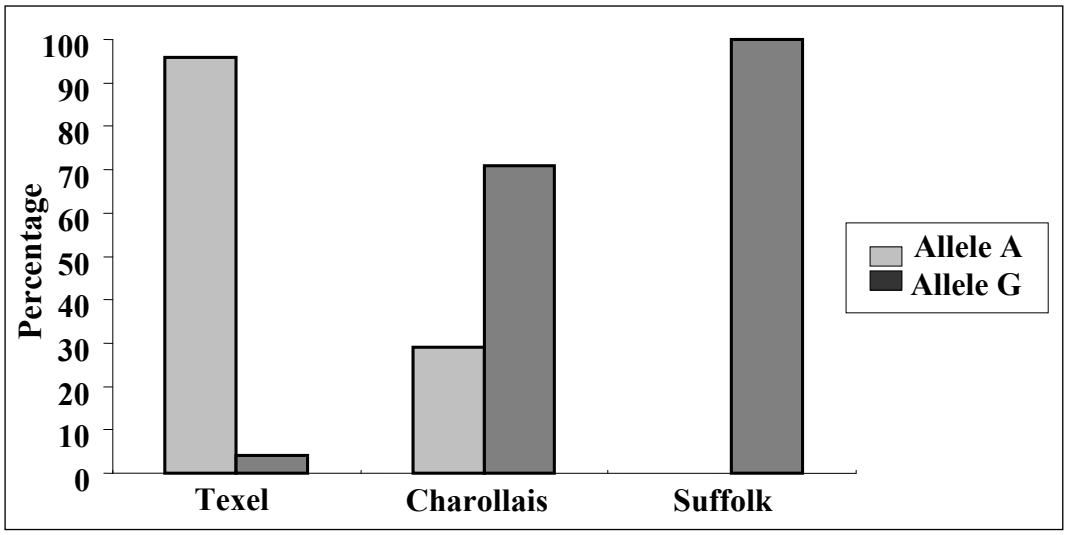
2.3 Results

2.3.1 Allele frequencies in British commercial sheep breeds

Allele frequencies at the *g+6723G>A* SNP locus in the three genotyped sheep breeds are shown in Figure 2.1. The A allele, associated with muscular hypertrophy, was absent in Suffolk sires that were genotyped and almost fixed in Texel sires. Both the *g+6723A* and the *g-2449C* alleles were segregating at intermediate frequencies (p=0.3) in the sample from the Charollais population. Fisher's exact test showed that

the association between the alleles of the two SNP loci was non-random (P-value=10^{-6}). The correlation between the alleles at the two SNP loci, r, and its square, r^2, were 0.95 and 0.90, respectively.

Figure 2.1 Allele frequencies at the *GDF8 g+6723G>A* SNP locus in British commercial sheep. Frequencies are based on locus genotypes of Texel and Suffolk sires and of Charollais lambs.



2.3.2 Association analyses of GDF8 SNP effects on traits in the Charollais breed

Means, standard deviations and ranges of the traits studied in the Charollais lambs are shown in Table 2.2. Direct genetic, maternal genetic heritabilities and the proportion of phenotypic variance due to common environment (litter) for muscle and fat depth are shown in Table 2.3. The direct genetic heritabilities of muscle and fat depth were moderate, whereas maternal genetic heritabilities for both traits were low. Common environment contributions to the phenotypic variance for muscle and fat depth were large, yet smaller than those of the direct genetic effects (Table 2.3).

Table 2.2 Trait means, ranges and phenotypic standard deviations

Trait	Mean	Range	sd[1]
Live weight (kg)	51.42	15.50-97.00	6.67
Muscle depth LV3[2] (mm)	28.31	12.00-45.00	2.91
Sqrt(Fat depth)[3] LV3[2] ($mm^{1/2}$)	1.88	0.10-3.92	0.40

[1]Phenotypic standard deviations after adjusting for fixed effects

[2]LV3 = 3rd Lumbar Vertebra

[3]Sqrt(Fat depth) refers to the square-root transformed trait. Raw fat depth data had a mean of 3.78 mm and a range of 0.01-15.33 mm

Table 2.3 Estimated trait variance ratios (variance component/phenotypic variance) for each random effect fitted in the selected REML model

Trait	Variance component		Variance ratio (±SE)
Muscle Depth	Animal	(Direct genetic effect, i.e. heritability)	0.29±0.011
	Dam	(Maternal genetic effect)	0.03±0.005
	Litter	(Common environmental effect)	0.22±0.007
Sqrt(Fat Depth)[1]	Animal	(Direct genetic effect, i.e. heritability)	0.31±0.011
	Dam	(Maternal genetic effect)	0.04±0.005
	Litter	(Common environmental effect)	0.28±0.007

[1]Sqrt(Fat depth) refers to the square-root transformed trait

Mixed model association analyses of the two *GDF8* SNPs were performed for muscle depth, fat depth, and live weight. None of the analyses showed significant association of genotype at either SNP locus with live weight. When the genotypic data for either or both SNPs were analysed using a sire model, significant effects were seen for both muscle ($P<0.01$) and fat depth ($P<0.05$) (results not shown). When the complete dataset was used, and an animal model was fitted, the significant fat depth effect of the SNP genotypes at either locus disappeared, but the overall SNP genotype effects on muscle depth remained. The overall *g+6723G>A* SNP genotype effect on muscle depth had an F-ratio of 8.05 ((2, ~303); $P<0.001$) and the *g-2449G>C* SNP effect on the trait had an F-ratio of 7.78 ((2, ~303); $P<0.001$) (see

Table 2.5 for the estimated allelic effects of each SNP). Genotype class contrasts for each SNP are discussed below. When either SNP genotype was fitted as a fixed effect after the other one was already included in the model, the second SNP fitted did not have a significant effect on muscle depth.

For each SNP, the significance of pairwise contrasts between the predicted trait values of the three genotype classes was determined using a two sample t-test (Table 2.4). Significant differences were found between the AA and GG, and the AA and AG genotype pairs at the *g+6723G>A* SNP locus for muscle depth, but not between GG and AG. For the g-*2449G>C* SNP locus, significant differences were detected when the CC genotype value was contrasted to either one of the GG and CG genotypes, whereas no significant differences were identified between the GG and CG values. Genotype class effects on square-root transformed fat depth were generally not significant, except for a marginally significant difference detected between the CC and GG genotypes at the g-2449G>C SNP (P~0.05).

Estimations using the predicted trait values for each SNP genotype class (Table 2.5) showed that the *g+6723A* allele had an additive effect of 1.20(±0.30) mm on muscle depth. The dominance effect of the A allele at the locus was negative, viz. -0.73 (±0.36) mm. For fat depth, the *g+6723A* allele was related to a small but non-significant decrease in fat depth (additive value=-0.065(±0.040) $mm^{1/2}$). The additive effect of the g-2449C allele on muscle depth was 1.00(±0.25) mm, and the dominance effect was -0.45(±0.33) mm.

Based on the estimated allelic effects and the allele frequencies observed in the sample at each locus, the proportion of additive genetic variance attributed to the SNP genotype at each SNP locus was determined for both muscle and fat depth

(Table 2.5). The *g+6723G>A* SNP genotype explained 14% of the additive genetic variance for muscle depth, and 2.1% for fat depth. The *g-2449G>C SNP* genotype accounted for 11% of additive genetic variance for muscle depth and for 1.5% for fat depth.

Table 2.4 Test statistics for contrasts between SNP genotype class values for muscle and fat depth using the animal model

Trait	Effect	Contrast	T-statistic[1]	P-value
Muscle Depth	*g+6723G>A*	A:A vs G:A	3.32 (168 df)	<0.002
		A:A vs G:G	4.01 (193 df)	<0.002
		G:A vs G:G	1.44 (309 df)	<0.20
	g-2449G>C	C:C vs G:C	2.96 (179 df)	<0.01
		C:C vs G:G	3.94 (193 df)	<0.001
		G:C vs G:G	1.72 (298 df)	<0.10
Sqrt(Fat Depth)[2]	*g+6723G>A*	A:A vs G:A	1.35 (168 df)	<0.10
		A:A vs G:G	1.62 (193 df)	<0.10
		G:A vs G:G	0.57 (309 df)	<0.5
	g-2449G>C	C:C vs G:C	1.88 (179 df)	<0.10
		C:C vs G:G	1.95 (193 df)	~0.05
		G:C vs G:G	0.23 (298 df)	<0.80

[1]The degrees of freedom for the test given in brackets are only approximate

[2]Sqrt(Fat depth) refers to the square-root transformed trait

Table 2.5 SNP allelic effects and percentage of additive genetic variance explained by *GDF8* SNP genotypes

Trait	SNP	a ± SE[1]	d ± SE[2]	Percentage of genetic variance due to SNP[3]
Muscle Depth (mm)	*g+6723G>A*	1.20±0.30	-0.73±0.36	14
	g-2449G>C	1.00±0.25	-0.45±0.33	11
Sqrt (Fat Depth)[4] ($mm^{1/2}$)	*g+6723G>A*	-0.065±0.040	0.040±0.049	2.1
	g-2449G>C	-0.066±0.034	0.056±0.044	1.5

[1]Negative additive genetic effect (a<0) indicates A (or C) allele decreased the trait

[2]If a>0 and d<0 (or a<0 and d>0), the A (or C) allele is partially recessive

[3]Estimated using allele frequencies observed in sample (p=0.3 for A or C allele)

[4]Sqrt(Fat depth) refers to the square-root transformed trait

Parent of origin effects, tested by contrasting the two heterozygote classes at each SNP locus, were generally not significant. An effect which approached significance was detected with the sire model association analysis of each SNP haplotype with muscle depth (heterozygote contrast = 2.82(±1.53) mm; P-value=0.07). However, the analyses were limited by the fact that in the sample only five heterozygous animals for the *g+6723G>A* SNP and only ten for the *g-2449G>C* SNP had inherited the minor allele from their dam.

2.4 Discussion

This study examined the presence of two *GDF8* single nucleotide polymorphisms in British commercial sheep breeds. It is the first published report of significant association of these polymorphisms with muscle depth in commercial Charollais sheep. In addition, an extensive study of the allelic and genotype effects on muscle depth was performed, and the mode of SNP action was determined for this breed, with the effect appearing to be partially recessive, at least for the trait measured. Finally, in accordance with the strategy outlined by Ron & Weller (2007), our analyses provided further statistical validation for the *g+6723G>A* SNP being a QTN for muscularity in sheep, as previously proposed by Clop *et al.* (2006).

The first objective of this study was to examine whether the same genetic mutation(s) were responsible for the similar phenotypes attributed to OAR2 QTL in different breeds (Walling *et al.* 2004; McRae *et al.* 2005). More specifically, we sought to determine whether the *g+6723A* and *g-2449C* alleles of the two *GDF8* SNPs associated with increased muscularity in Belgian Texel rams, were present in

British terminal sire breeds. Since the favourable alleles were absent in the Suffolk animals sampled and nearly fixed in the Texel breed (Figure 2.1 for the *g+6723G>A* SNP results), other OAR2 loci must be responsible for the QTL detected for muscle and fat traits in these breeds. Fixation of the A allele at the 3'-UTR *g+6723G>A* SNP locus was previously found for Belgian Texel rams (Clop *et al.* 2006). The Charollais is the only breed, apart from the Australian White Suffolk (Kijas *et al.* 2007), in which these *GDF8* SNPs have been found to segregate, with the favourable allele of each SNP having an intermediate frequency (p=0.3) in the genotyped animals.

A mixed model in which animal, dam and litter were fitted as random effects revealed highly significant association of SNP genotype at both *GDF8* SNP loci with muscle depth in our lamb population. Estimates for the direct genetic, maternal genetic heritabilities and the proportion of variance due to common environment (litter) were consistent with published values of these genetic parameters for these traits in Charollais and other sheep breeds (Ap Dewi *et al*. 2002; Jones *et al.* 2004; Safari *et al.* 2005). In these analyses the age at scanning (in days) was fitted as a covariate for each trait to adjust for systematic trends in trait values associated with age *per se*. Alternatively, fat and muscle depths may be corrected for live weight at scanning or live weight together with age at scanning, to remove differences in these traits attributable to animal size. When these analyses were performed, i.e. fitting live weight as a covariate or live weight and age as covariates, the SNP genotype association results and the magnitude and direction of the estimated additive and dominance effects were essentially the same as those seen when the data were corrected for age at scanning. Including both live weight and age as covariates in the

analysis of muscle depth resulted in the predicted mean trait value for the heterozygote at each SNP locus being significantly different from the trait means of both homozygotes. Prior to the association analysis, the fat depth data were subjected to square-root transformation to ensure that a normally distributed trait would be analysed. All results have been presented on this transformed scale. Back-transformation of the genotype values yielded predicted mean fat depths of 2.80, 3.16, and 3.25 mm for the AA, AG, and GG genotypes, respectively.

The two SNPs, which are about 9kb apart in the *GDF8* region (Clop *et al*. 2006), exhibit substantial LD (r^2=0.90). Thus, if the real functional effect on muscle depth arises from the *g+6723A* allele, as proposed by Clop *et al.* (2006), then the detected effect of the *g-2449C* allele is probably due to its LD with the causative variant. Additive and dominance values for the *g-2449G>C* SNP were comparable in size and in sign with those for the *g+6723G>A* SNP, as would be expected from the strong LD between the two SNPs.

The opposite signs of the allelic effects (positive additive and negative dominance value) supported a partially recessive nature of the causative *g+6723A* allele for muscle depth. This is in accordance with the previously detected partially recessive mode of action of the OAR2 QTL (later mapped to the *g+6723A* allele) on various traits related to muscularity in an F2 population of Belgian Texel x Romanov cross (Clop *et al.* 2006). In a recent study of the *g+6723A* locus in New Zealand Texel sheep, significant additive effects of the *g+6723A* allele on muscle and fat traits were also described (Johnson *et al.* 2009). Dominance effects on the traits were not significant in these analyses, although, overall, the trait changes associated with the *g+6723A* allele were greater in animals with two copies than in heterozygotes.

The proportion of additive genetic variance for muscle depth accounted for by SNP genotype at each *GDF8* SNP locus depends on the allele frequencies. The values presented in Table 2.5 for the *g+6723G>A* and the *g-2449G>C* SNP (14% and 11%) increase to 29% and 21% of the additive genetic variance, respectively, if allele frequencies of 0.5 are assumed. Due to the partially recessive mode of the A allele for muscle depth, the breeding and genotypic values for each genotype are also strongly affected by the allele frequencies at the locus. Further, it can be shown that the additive and total genetic variance explained by the SNP genotype will maximise when the frequency of the *g+6723A* allele is 0.70 and 0.68, respectively. The maximum additive genetic variance that can be attributed to the SNP genotype at this frequency corresponds to 38% of the total additive variance of muscle depth. Thus, based on both the estimated magnitude of the SNP effect on muscle depth and the amount of genetic variance of the trait that it explains, genetic selection for the *g+6723G>A* QTN is of particular economic importance for the Charollais breed.

The detection of partially recessive action of the *g+6723A* allele on the muscle phenotype raises the issue of whether a recessive mode of allelic expression is indeed plausible. The proposed molecular mechanism of miRNA-mediated translational inhibition of myostatin by which the *g+6723A* allele leads to increased muscle (Clop *et al.* 2006) does accommodate a partially recessive action of the mutation on myostatin expression. According to this mechanism, in a heterozygous animal, translational inhibition of the mutant mRNA would lead to about one third of the normal amount of myostatin produced (Clop *et al.* 2006), whereas the wild type *GDF8* mRNA would produce normal amounts of protein. Thus, it is likely that enough active myostatin would be produced to facilitate myostatin-mediated

regulation of muscle development. This would result in the observed partially recessive effect on the phenotype.

It should be noted that the overall mode of action and magnitude of the observed effect of the A allele on phenotype would probably depend on the genetic background of a particular animal population. Indeed, the influence of background genetics on the effects of *GDF8* mutations on phenotypes has been shown in cattle (Short *et al.* 2002) and mice (Bünger *et al.* 2004; Rehfeldt *et al.* 2005). Additionally, the effect of overexpression of other molecules interacting with myostatin and other growth factors on muscle development was recently shown in mice (Lee 2007). The magnitude and perhaps even the mode of action of the *GDF8* mutational effects on phenotype are probably affected by the genetic background of the breed in which the mutation is seen.

In Great Britain, the frequency of the favourable A allele of the *g+6723G>A* SNP in the Charollais breed has probably increased somewhat due to mass selection on muscle traits. Yet, because of its partially recessive action on muscle phenotype, the rate of genetic progress for the trait depends heavily not just on the allele frequency but also on the proportion of homozygote animals for the A allele in the population. Consequently, marker assisted selection (MAS) for this SNP could be of substantial benefit. Given the above, it actually may be even more advantageous to apply MAS for a (partially) recessive than for a dominant allele. This is because, if mass selection (and not MAS) is applied, heterozygous animals would not be chosen since the dominant allele, which is unfavourable for the trait, would not allow the expression of the recessive allelic effects. Overall, for maximum gain from MAS, a comprehensive strategy for nucleotide assisted selection that takes into account a

SNP's mode of action on the traits of interest, and the effects of the genetic background of the breed or population, should be employed.

2.5 Conclusions

To assess whether the same mutation(s) resulted in similar phenotypes attributed to OAR2 QTL in Suffolk, Texel, and Charollais breeds, rams from British commercial flocks were genotyped for two SNPs located in the *GDF8* region of OAR2, previously detected in progeny of Belgian Texel rams exhibiting muscular hypertrophy. The *g-2449C* and *g+6723A* alleles were absent in the Suffolk sires sampled, almost fixed in the Texel and segregating in the Charollais sires. Mixed model association analyses using SNP data on 338 Charollais lambs revealed that both SNPs had a significant association with muscle depth ($P<0.001$). The SNPs were segregating at intermediate frequencies ($p=0.3$) and exhibited strong linkage disequilibrium ($r^2=0.90$). Animals with the AA genotype at the *g+6723G>A* SNP locus had significantly greater muscle depth than those with either the GG or the AG genotype ($P<0.002$), with the A allele, the likely causative mutation, having an additive effect of 1.20(±0.30) mm and a dominance effect of -0.73(±0.36) mm. Based on estimated allelic effects and sample allele frequencies, the *g+6723G>A* SNP explained 14% of the additive genetic variance of muscle depth. The maximum genetic variance for the trait attributed to the locus would be attained at a *g+6723A* allele frequency of 0.7 and would equal 38%. These findings indicate that marker assisted selection using these two *GDF8* SNPs would benefit the Charollais breed.

CHAPTER 3

A growth model approach for dissecting quantitative trait loci (QTL) affecting growth traits in Scottish Blackface sheep

3.1 Introduction

Growth is an economically important trait for the sheep industry as it is directly related to meat production. Production of faster-growing lambs would be highly beneficial for producers since, apart from the fact that higher lamb weight would mean greater revenues, it would result in enhanced feed conversion efficiency. This would lead to various benefits, including lower production costs, higher product yields, less nitrogenous-waste excretion to the environment and decreased grazing pressure (Cockett *et al.* 2005). Genetic selection is a valuable approach for achieving improved lamb growth. For more effective genetic selection on growth, it is advantageous to identify the genetic loci that influence growth in terms of body weight and weight gain of each animal.

Several studies in livestock (although few in sheep) have reported quantitative trait loci (QTL) associated with growth traits in terms of average daily gain, weight at a specific age, and days to reach a particular weight (e.g. Stone *et al.* 1999 for cattle; Nagamine *et al.* 2003 and Stearns *et al.* 2005 for pigs). The majority of QTL mapping studies have used univariate approaches to detect QTL, treating

weights recorded at a particular growth point as separate traits. This is despite the fact that live weights across time comprise a longitudinal trait that is a function of several physiological processes and a composite of phenotypes recorded over time. Thus, strong genetic correlations exist among live weights at different ages (Corva & Medrano 2001, Riggio *et al.* 2008), although patterns of correlations often suggest additional complexity. For example, using sheep data Riggio *et al.* (2008) showed that inter-age genetic correlations for live weight, whilst strongly positive, are often different from unity, with the correlation decreasing as the time between the weight measurements increases. Thus, it is likely that distinct loci act on live weights at different growth stages. For the detection of QTL that are associated with growth or live weight, it would be beneficial to simultaneously analyze multiple measurements and take account of the correlation structure of measurements across time.

Fitting growth models on body weight data from different time-points and extracting the relevant growth parameters provides a way to combine phenotypic information from multiple measurements into a few variables in a biologically meaningful manner. This approach has been previously applied in livestock (Lopez *et al.* 2000, several species; Schinckel *et al.* 2004, pigs; Lambe *et al.* 2006, sheep). Flexible sigmoid curves, such as the logistic, the Brody, Bertalanffy, Gompertz and Richards curves, often represent the best-fitting growth models and are extensively described in the literature for livestock (e.g. Pittroff *et al.* 2008). Many growth curve variables relevant to genetic studies may be derived from such models, describing growth rates and live weights as well as features such as maximum growth rate, and the age at which it is predicted to occur, and mature weight.

In sheep, there is little published information on the genetic control of growth curve variables and the different stages of growth. A few studies have investigated the polygenic components of growth curve parameters using growth models describing weight and growth rate (as the derivative of the weight function) as a function of time (Lewis *et al.* 2002; Lambe *et al.* 2006) or random regression methodology (Lewis & Brotherstone 2002; Lambe *et al.* 2006). These studies indicated that growth variables are indeed heritable and that genetic differences seem to exist among growth parameters of various breeds. Moreover, the study conducted by Lambe *et al.* (2006) suggested that early growth rate is a different genetic trait to later growth rate. All these postulations support a hypothesis that growth curve variables are under genetic regulation and that they may constitute separate aspects of the complex longitudinal trait of growth. Validation of this assertion requires a more detailed description of the genetic control of growth. To this end, it would be informative to dissect the genetic loci that underlie growth curve predictors.

To date, no QTL study on growth curve parameters has been performed in sheep or any other livestock species, although a Bayesian procedure for QTL detection using prior information obtained from a growth model was applied in pigs (Varona *et al.* 2005). The main objective of this study was to identify and describe QTL in Scottish Blackface sheep that directly influence longitudinal live weights and growth as a process, using descriptors of growth derived from fitted growth curves. We also aimed to examine whether, for particular growth traits, the effects of different QTL were constant over time or changed as the animals grew. If the latter were the case, we were interested in quantifying how the QTL effects changed over time.

3.2 Materials and Methods

3.2.1 Animals and traits measured

The population studied has been previously described in detail (Davies *et al.* 2006; Karamichou *et al.* 2006). In brief, the population consisted of 830 Scottish Blackface lambs from nine half-sib families, with progeny per family ranging from 34 to 154 individuals. The animals were bred over a 3-year period (2001–2003). Standard records (such as parentage, day of birth, sex) and weight measurements at birth and at four-week intervals after birth (up to 24 weeks) were collected, ranging from 830 to 691 records per time point. The distribution of live weights across age is given in Figure 3.1. The complete pedigree for this population contained 4866 animals, with records dating back to 1986.

3.2.2 Genotyping and linkage map construction

Lambs were genotyped for informative microsatellite markers, i.e. markers that were heterozygous in their sire, on chromosomes 1, 2, 3, 5, 14, 18, 20 and 21, as detailed in Davies *et al.* (2006) and Karamichou *et al.* (2006). A linkage map was constructed for the markers on each chromosome using the "build", "all", and "flips" options of CriMap version 2.4 (Green *et al.* 1990). The marker order with the highest likelihood was selected in order to construct the consensus linkage map for each chromosome that was subsequently used in all QTL analyses (Appendix 3.1). The linkage maps constructed were in close agreement with those from other mapping studies (Maddox *et al.* 2001); http://rubens.its.unimelb.edu.au/~jillm/jill.htm).

3.2.3 Phenotypic data treatment

Initially, live weights measured at birth and four-week intervals were treated as separate traits. Multiple regression analyses were performed on each phenotype (using R statistical package) in order to identify significant fixed effects. Fixed effects significant for all live weights were sex (2 levels), litter size reared (2 levels), age of dam (4 levels) and year by management group (i.e. field) (6 levels). The day of birth or the age at time of measurement was fitted as a covariate for each trait. These fixed effects and covariates were fitted in all subsequent regression analyses.

3.2.4 Growth model choice

Five growth functions were fitted to live weight measurements from all 788 animals for which five or more data points were available, using nonlinear regression in SAS (release 9.1, SAS Inst., Cary, NC). The nonlinear growth functions fitted were the generalized Michaelis-Menten (GMM) (Lopez *et al.* 2000), the Gompertz, the logistic, the Richards and the exponential models. The Gompertz, logistic and Richards functions have been described and applied by Renne *et al.* (2003) and Lambe *et al.* (2006) and are special cases of a more general model (Turner *et al.* 1976). The reparameterised version of the exponential model has been explained by Bünger & Herrendörfer (1994) and employed by Lambe *et al.* (2006). The formulas and parameter details for the above growth functions are given in Table 3.1. These parameterisations of growth models were chosen in order to study parameters with direct biological interpretation as explained previously in Renne *et al.* (2003).

Table 3.1 Growth model equations

Growth function	Parameters	y(t)[1]
Generalized Michaelis-Menten[2]	W_o, W_f, K, c	$(W_oK^c+W_ft^c)/(K^c+t^c)$
Gompertz[3]	A, B, C	$Ae^{\{-e^{[Be(C-t)/A]}\}}$
Logistic[3]	A, B, C	$A\{1+e^{[4B(C-t)/A]}\}^{-1}$
Richards[4]	A, B, C, D	$A\{1+D\,e^{\{[B(C-t)(D+1)^{1+1/D}]/A\}}\}^{1/D}$
Exponential[5]	A, B_E, C_E	$A-(A-C_E)e^{[B_Et/(A-C_E)]}$

[1]y(t) is the live weight at time t

[2]Described by Lopez *et al.* 2000

[3]Reparameterisations described by Renne *et al.* 2003

[4]Reparameterised by E. Schönfelder, Institut für angewandte Tierhygiene, Eberswalde, Germany (Lambe *et al.* 2006)

[5]Reparameterised as suggested by Bünger and Herrendörfer 1994

Growth model choice followed a similar procedure as described in detail by Pittroff *et al.* (2008). Specifically, each model was first fitted to the dataset as a whole, i.e. one curve fitted to all the data as shown in Figure 3.1, and three models were immediately rejected: (i) the logistic model provided a poor fit to the live weight data and it generally would not converge unless one of the model parameters was fixed *a priori*; (ii) the exponential model fitted the live weight data overall but no estimate for parameter C_E was obtainable; (iii) the Richards model converged for the live weight data, and provided an overall good fit, as assessed by the Residual Mean Square (RMSQ), but resulted in a high correlation between parameter estimates of C and D (r^2=0.98). However, the GMM and Gompertz models for growth provided a good overall fit to the data, low correlations were observed between the estimated parameters fitted and the RMSQ were comparable for the two models. Thus, the Gompertz and GMM models were used to model live weights for

each animal separately, using both the NLIN procedure of SAS (release 9.1, SAS Inst., Cary, NC) and the nonlinear modelling procedure of JMP (release 7, SAS Inst., Cary, NC). The JMP procedure provided graphical representation of the fit while the model fitting iterations were running and thus allowed immediate examination of prediction bias, i.e. systematic deviations between the observed and fitted values.

Although the GMM model fitted the whole dataset well, it provided a poorer fit when applied to individual animals, resulting in negative predicted birth weights for some animals and implausible values for mature weight for 80 out of 788 animals, even when convergence was achieved. This characteristic was also observed by Pittroff *et al.* (2008). Therefore, the Gompertz growth model was chosen as it converged for each of the 788 animals in the dataset and had low apparent prediction bias.

The Gompertz model was then used to predict live weights and growth rates at weekly intervals from birth up to 24 weeks of age for all 788 lambs. Additionally, maximum growth rate and time at maximum growth was estimated for each lamb. The following equations were used:

$y(t) = Ae^{\{-e^{[Be(C-t)/A]}\}}$ and $dy/dt = [(B/A)e]y(t)\ln[A/y(t)]$, where: y(t) = live weight at time t; A = estimated final body weight, kg; B = maximum growth rate (average daily gain), kg/d; C = age at maximum growth rate, d; A/e= live weight at maximum growth, kg.

3.2.5 Half-sib QTL regression model

QTL analyses were conducted using a univariate multi-marker approach for interval mapping in half-sib families, as described by Knott *et al.* (1996) and applied by the web-based software package QTL express (Seaton *et al.* 2002). The probability of inheriting a particular sire allele was calculated at 1cM intervals for each offspring, conditional on the marker genotypes of the individual and its sire and on the sire's linkage phase. Subsequently, the trait phenotype was regressed on the conditional probability of the offspring genotypic inheritance for a given position. For each regression, an F-ratio of the full model including the significant fixed effects (sex, litter size reared, age of dam, and year by group), covariate (day of birth or age at measurement) and the inheritance probabilities versus the same model without the inheritance probabilities was calculated. The chromosomal location with the largest F ratio was taken to be the best estimated position for a QTL for each trait. In addition, the within-sire substitution effects for each sire family were obtained from the analysis. An estimate of the overall QTL effect was obtained by calculating the average of the absolute values of the QTL allelic substitution effects across families for which the QTL was significant. The proportion of phenotypic variance, corrected for all fitted fixed effects and covariates (Vp), that was explained by the QTL for each trait was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$ (Knott *et al.* 1996), where "full" is the model with the QTL effect fitted and "reduced" is the model without the QTL effect.

3.2.6 Significance thresholds

Chromosome-wide empirical threshold values were determined for the test statistics obtained from the regression analysis at $\alpha = 0.05$ and 0.01 by applying one thousand chromosome-wide permutations for each trait (Churchill & Doerge 1994). Threshold values varied between chromosomes depending on their length and marker content. For all chromosomes, the nominal threshold for significance was determined for a single test for QTL detection, using the F-ratio (P-value<0.05) for the model including the QTL as a fixed effect [(9, 711-750) df]. Thus the nominal F-ratio was set to F-ratio>1.89.

3.2.7 Traits analysed

Prior to QTL analysis, the distribution of each of the extracted model variables (A, B, C) was examined and extreme outliers, deviating more than three standard deviations from the mean, were removed. The distribution of the C variable was skewed, and, therefore, the C estimates were transformed using a natural log transformation. All other parameters were analysed without transformation since they appeared normally distributed.

Live weight at birth and each of four-week intervals was subjected to univariate interval QTL mapping for each chromosome using data and marker genotype information at each chromosome, as described above. Additionally, each of the Gompertz model parameters (A, B, ln(C)), predicted growth rates and live weights at weekly intervals and at maximum growth (point of inflection) were analysed for QTL detection.

3.3 Results

3.3.1 Gompertz growth model description

The growth curve resulting from fitting the Gompertz model to the combined live weight data from all animals is shown in Figure 3.1. For each model fitted which converged, the residual mean squares (RMSQ) are shown in Table 3.2 along with F-ratios for the fit of the model. The means for Gompertz model parameter estimates (averaged across the predicted values for each of the 788 animals), their standard errors, and asymptotic (i.e approximate) 95% confidence intervals are given in Table 3.3. Using the Gompertz model, live weights were predicted for each of the 788 animals at each time point with actual live weight measurements, and no significant differences were found using a t-test (results not shown) between the predicted and observed values at all ages examined.

Table 3.2 Least square statistics for each non-linear regression model that was fitted to the entire live weight dataset[1]

Model	RMSQ	F-ratio	Numerator df[2]	P-value
Gompertz	1569	68769	3	<0.0001
Generalized Michaelis-Menten	1538	52648	4	<0.0001
Richards	1551	52173	4	<0.0001
Exponential	12085	10975	3	<0.0001

[1] Model equations and the parameters fitted for each model are described in Table 1

[2] Numerator degrees of freedom (df) corresponded to the number of parameters fitted for each model. Denominator df corresponded to the residual df and ranged from 5549-5551 across models.

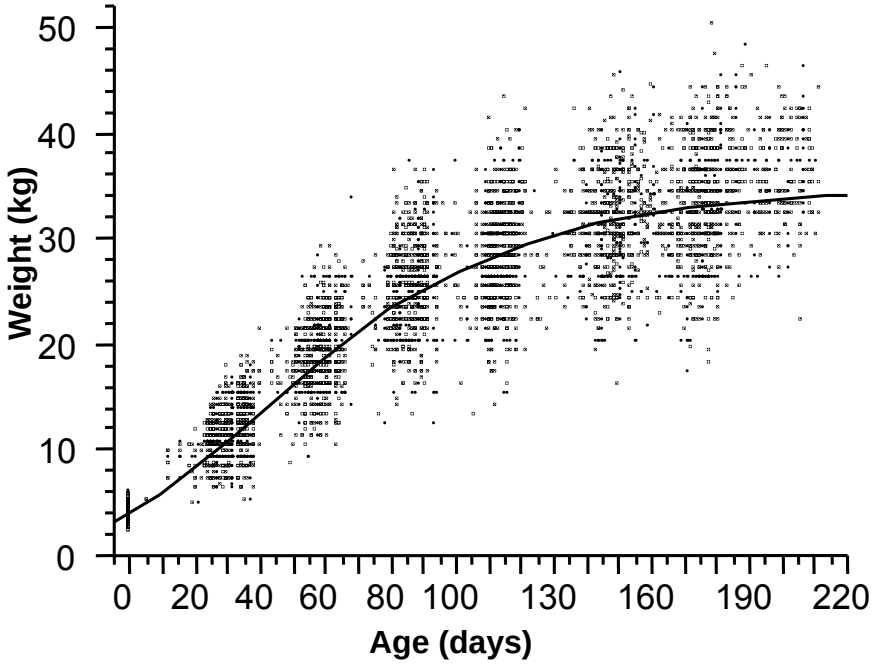


Figure 3.1 Distribution of live weights of Scottish Blackface lambs across age (dots) and the average growth curve obtained after fitting the Gompertz model to the weight data (solid line).

Table 3.3 Estimated means and standard errors for Gompertz model parameters and weight at point of inflection[1]

Parameter	Sample mean	Approx. SEM[2]	Approx. 95% Confidence Interval
A (kg)	35.1	0.19	34.72-35.47
B (kg/days)	0.284	0.002	0.280-0.288
C (days)	36.9	0.32	36.27-37.54
ln(C) (ln(days))	3.58	0.009	3.559-3.594
Weight at point of inflection (kg)	12.9	0.08	12.77-13.07

[1]The average of individual lamb means, predicted after fitting the Gompertz model to live weight measurements for each animal, was estimated for each parameter

[2]SEM=Standard Error of the Mean

3.3.2 QTL results

All QTL for observed live weight whose significance exceeded the 5% chromosome-wide threshold are reported in Table 3.4, along with significant QTL for the Gompertz function parameters. The live weight QTL were detected on the complete dataset and no changes in the location or significance of these QTL were observed when the dataset of observed live weights was reduced to the 788 animals included in the Gompertz model procedure (results not shown).

For predicted live weight and growth rates the detected QTL tended to be significant over a period of one or more weeks, and F-ratio trajectories for these QTL are shown in Figures 3.2, 3.3, 3.4 and 3.5, for chromosomes 20, 14, 3 and 18, respectively. A summary of these QTL is given in Table 3.5, giving results for the time points at which the QTL were of maximum significance. Chromosome-wide significance thresholds varied marginally over time, and plotted in each figure is the maximum observed threshold on each chromosome across age for the relevant traits. The F statistic presented in the trajectories in Figures 3.2-3.5 is for the best estimated position at each time point. This position was consistently within a maximum difference of 1-3 cM between the most distant time points on a given trajectory. More importantly, the QTL resided in the same marker interval and was detected in the same families for all plots (connected points) for predicted growth rate or weight.

Table 3.4 Summary of significant QTL for observed live weights and growth curve functions from across-family univariate QTL analyses

Trait	Chr.	QTL Pos. (cM)[1]	Marker Interval	F ratio[2]	Families Sign.[3]	% Vp explained by QTL[4]	QTL Effect[5]
Birth Weight (kg)	14	82	BMS833	4.13 (2.76, 3.38)	1, 6	14	0.912
8wk-weight (kg)	14	110	ILSTS002-LSCV30	2.8 (2.47, 3.00)	3, 8	9	2.51
16wk-weight (kg)	20	61	DRB1-TGLA387	2.53 (2.47, 3.03)	4, 6	8	4.63
Growth rate at point of inflection (B) (kg/day)	20	60	DRB1-TGLA387	2.92 (2.41, 2.92)	5, 6	9	0.054
Age at point of inflection (C or ln C) (days or ln days)	14	98	ILSTS002-LSCV30	3.18 (2.37, 2.81)	8	11	ln(C)=0.166 C=1.18
	5	123	MCM527-CSRD2134	2.41 (2.33, 2.78)	4, 8	7	ln(C)=0.166 C=1.18

[1]QTL position is defined relative to the first marker present in the genetic map for each chromosome; first marker positioned at 0 cM.

[2]Chromosome-wide F-statistics for P-values < 0.05 and < 0.01 (as determined by permutation testing) are given in parenthesis.

[3]Families within which a QTL effect was deemed significant using a t-test when a half-sib QTL regression model was fitted across families.

[4]Vp refers to the phenotypic variance for each trait, after correction for all fitted fixed effects and covariates. The proportion of Vp due to the QTL was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$, where "full" is the model with the QTL effect fitted and "reduced" is the model without the QTL effect (Knott *et al.* 1996).

[5]QTL allelic substitution effect, determined as the average of the estimated absolute values across families for which the QTL was significant. It corresponds to the difference in trait values between the two QTL alleles that can be inherited from a sire heterozygous for the QTL.

The significance trajectories for a QTL on chromosome 20 for predicted growth rates and weights across age are given in Figure 3.2. The chromosome 20 QTL for growth rate became significant (at the chromosome-wide level) at three weeks, its significance maximised at six weeks and it retained at least nominal significance up to nine weeks (Table 3.5). A QTL located in the same marker interval on chromosome 20 and in the same families (5 and 6) was also highly significant for maximum growth rate (parameter B of the Gompertz model; Table 3.4). This QTL was apparent for predicted weight at a later age, being nominally significant from eight to 17 weeks, with the highest F-ratio at 12 weeks (segregating in families 4 and 6). For observed live weight, this QTL (i.e. same position and segregating in the same two families; 4 and 6) was observed at 16-weeks (Table 3.4). For observed 12-week weight, QTL segregation was significant in one of the two families (family 6; t-test=2.85) but did not reach nominal significance across all nine families (F-ratio=1.79). In summary, a growth rate QTL on chromosome 20 was observed at and around the estimated age of maximum growth, which subsequently manifested itself as a live weight QTL.

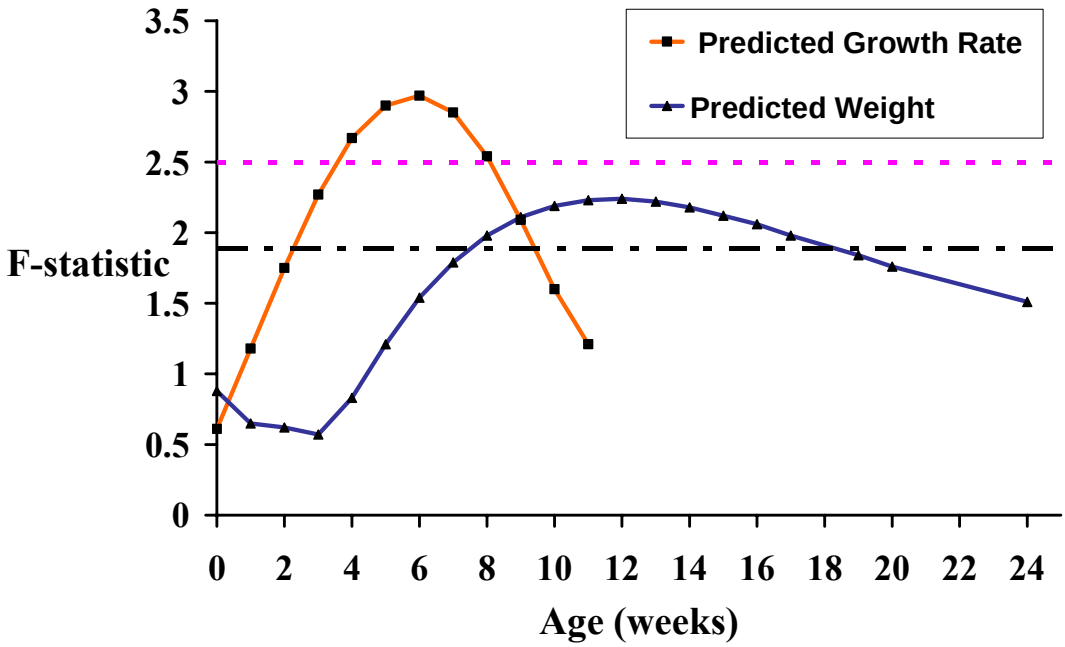


Figure 3.2 Across-age significance of QTL on chromosome 20 for live weights and growth rates, predicted using the Gompertz curve for weekly intervals. The chromosome-wide significance threshold (P-value<0.05; pink dashed line) was determined by permutation testing. The nominal threshold for significance (P-value<0.05; black dashed line) was estimated for a single test. The estimated QTL position and the QTL-segregating families for each trait are given in Table 3.5.

Table 3.5 Summary of significant QTL from across-family QTL analyses of predicted live weights and growth rates, at the time point when the significance was maximum

Trait	Chr.	QTL Pos. (cM)[1]	Marker Interval	F-ratio[2]	Sign. Range[3]	Families Sign.	QTL % of Vp[5]	QTL Effect[6]
1wk-weight (kg)	3	213	AGLA293-BL4/LYZ	3.46 (2.49, 2.86)	Birth to 4	2, 3, 8	12	0.696
12wk-weight(kg)	20	59	DRB1-TGLA387	2.24 (2.47, 2.93)	8-17	4, 6	6	3.19
1wk-growth rate (kg/day)	14	105	ILSTS002-LSCV30	2.19 (2.32, 2.98)	1-2	3, 8	6	0.034
6wk-growth rate (kg/day)	20	60	DRB1-TGLA387	2.97 (2.45, 2.96)	3-9	5, 6	10	0.051
10wk-growth rate (kg/day)	18	59	TGLA337-TGLA122	2.25 (2.23, 2.77)	8-12	3, 7	6	0.025
24wk-growth rate (kg/day)	14	99	ILSTS002-LSCV30	2.36 (2.37, 2.90)	17-28	8, 9	7	0.017

[1]QTL position is defined relative to the first marker present in the genetic map for each chromosome; first marker positioned at 0 cM.

[2]Chromosome-wide F-statistics for P-values < 0.05 and < 0.01 (determined by permutation testing) are given in parenthesis.

[3]Age range of significance given in weeks. Refers to chromosome-wide significance for all traits apart from 12wk-weight and 1wk-growth rate for which the age range of nominal significance is given.

[4]Families within which a QTL effect was deemed significant using a t-test when a half-sib QTL regression model was fitted across families.

[5]Vp refers to the phenotypic variance for each trait, after correction for all fitted fixed effects and covariate. The proportion of Vp due to the QTL was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$, where "full" is the model with the QTL effect fitted and "reduced" is the model without the QTL effect (Knott *et al.* 1996).

[6]QTL allelic substitution effect, determined as the average of the estimated absolute values across families for which the QTL was significant. It corresponds to the difference in trait values between the two QTL alleles that can be inherited from a sire heterozygous for the QTL.

Significant QTL were detected on chromosome 14 for observed birth weight and eight-week weight, although the families in which these QTL segregated differed for the two QTL. Although QTL for predicted weight at birth and eight weeks were not deemed significant across all families, a birth weight QTL segregated in family 6 and a QTL at eight weeks was significant in the same families (3 and 8) as for the observed eight-week QTL. Further, a nominal QTL on chromosome 14 was detected for growth rates at birth, weeks 1 and 2 in families 3 and 8; these being the same families and chromosomal location as seen for the eight-week weight QTL. Thus, the trend of F-ratio trajectories (Figure 3.3a) is the same as seen on chromosome 20, i.e. the QTL significance varies with time and its significance for growth rate occurs at an earlier age than that for live weight.

A further chromosome 14 QTL became nominally significant in families 8 and 9 for growth rate at 17 weeks and significant at 24 weeks (Figure 3.3b). This QTL may represent a separate locus associated with late growth; however it is possible that different allelic effects for the same QTL are being detected for early and late growth rate. This hypothesis is supported by the fact that in family 8 segregation for the QTL was statistically significant in both growth stages, with the estimated QTL effects for the two occasions being of opposite sign.

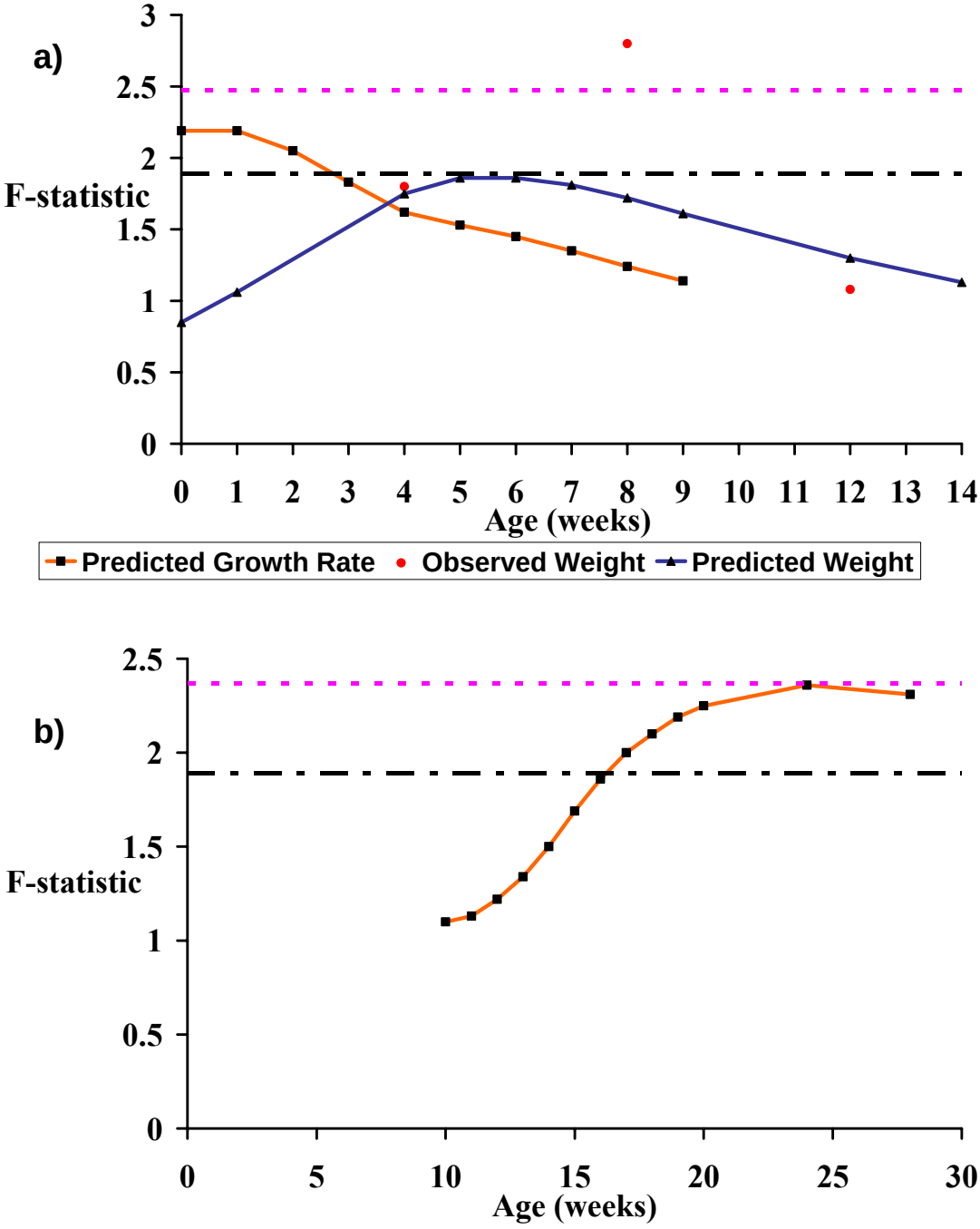


Figure 3.3 Across-age significance of **a)** early and **b)** late growth QTL on Chromosome 14. Weekly live weights and growth rates were predicted using the Gompertz curve. Chromosome-wide significance threshold (P<0.05; pink) was determined by permutation testing. Nominal threshold for significance (P<0.05; black) was estimated for a single test. The QTL position and the QTL-segregating families for each trait are given in Tables 3.4-3.5.

Two additional chromosomes yielded significant QTL for growth rate or live weight. The F-ratio trajectory across age of a QTL for live weight on chromosome 3 is shown in Figure 3.4. This QTL was highly significant for estimated weights at early ages (birth up to week 4) and significant within (but not across) the same three families (2, 3, 8) for observed weights at these ages. Yet, this QTL was not detected for predicted growth rates at any age point. A growth rate QTL on chromosome 18 was of nominal significance at eight weeks (Figure 3.5). The F-ratio for this QTL increased with age and became significant for growth rate at 10 weeks. The QTL significance remained nominal up to 12 weeks. No QTL was detected on this linkage group for observed or predicted live weights within the age range of our dataset.

Finally, two QTL were detected for the log-transformed age at point of inflection (population mean parameter C is 36.9±3.02 days; Table 3.3). Significant QTL were seen on chromosomes 14 and 5 (Table 3.4). No significant QTL were found for estimated final weight (parameter A).

Inspection of the estimated QTL effects for each observed or predicted growth trait (Tables 3.4 and 3.5) suggests that the size of the detected QTL effects, in terms of the proportion of phenotypic variance explained by the QTL at its maximum significance, remained roughly constant across QTL detected on different linkage groups. However, the absolute size of distinct QTL effects varied according to the trait means, i.e. live weight QTL effects increased with age, whereas the growth rate effects initially increased, then declined as growth rate slowed.

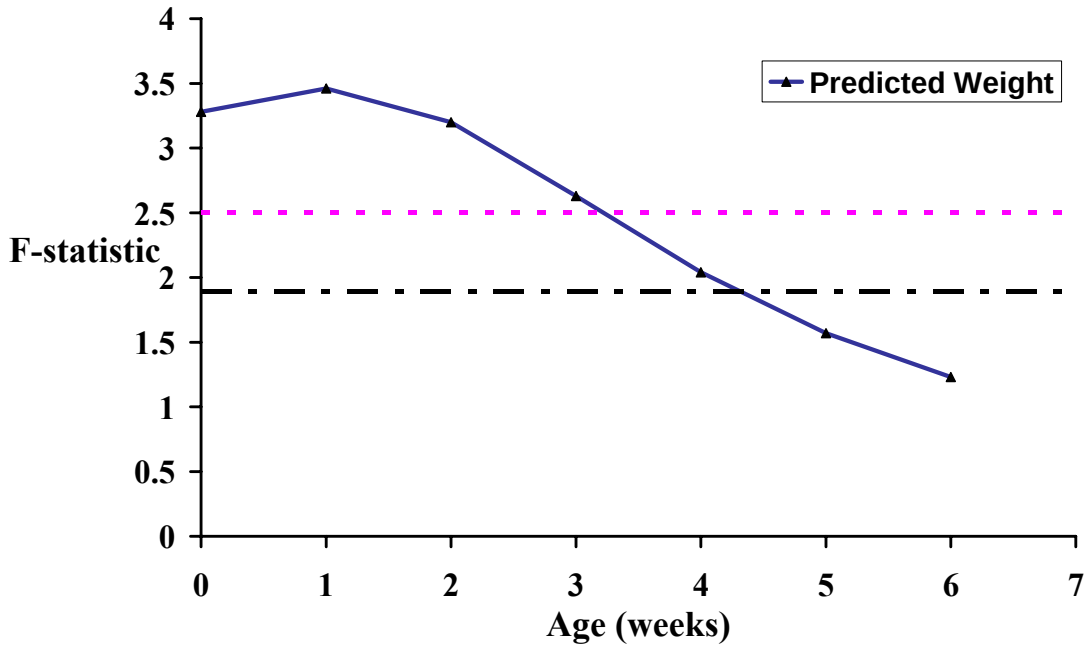


Figure 3.4 Across-age significance of QTL for predicted weight on chromosome 3. Live weight was predicted using the Gompertz curve for weekly intervals. The chromosome-wide significance threshold (P-value< 0.05; pink dashed line) was determined by permutation testing. The nominal threshold for significance (P-value<0.05; black dashed line) was estimated for a single test. The estimated QTL position and the QTL-segregating families are given in Table 3.5.

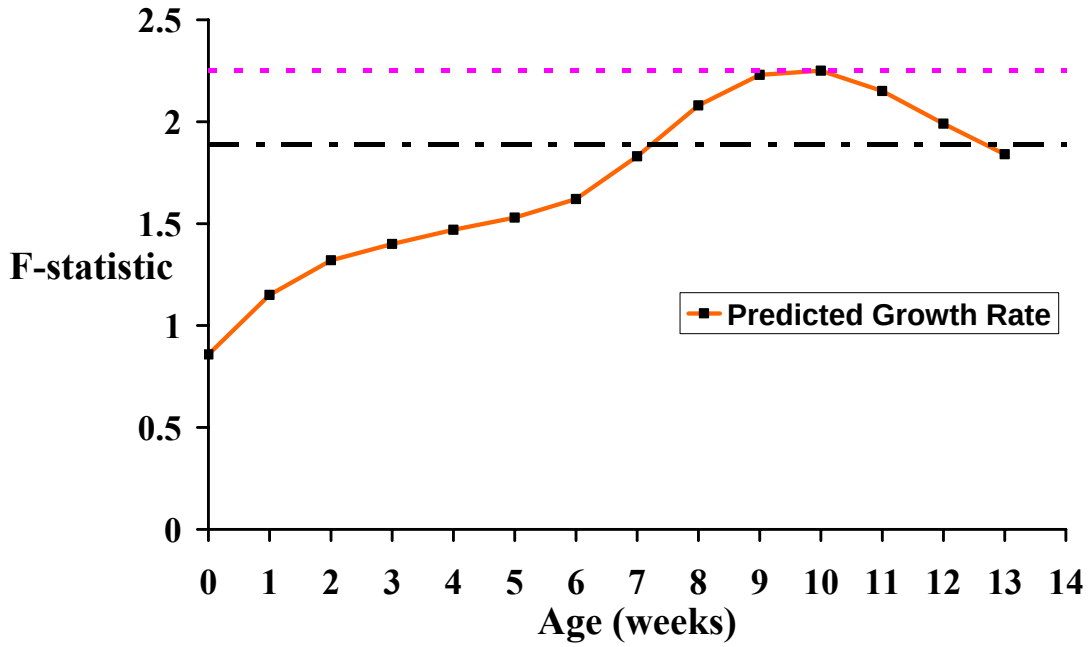


Figure 3.5 Across-age significance of QTL for growth rate on Chromosome 18. Growth rates were predicted using the Gompertz curve for weekly intervals. The chromosome-wide significance threshold (P-value<0.05; pink dashed line) was determined by permutation testing. The nominal threshold for significance (P-value<0.05; black dashed line) was estimated for a single test. The estimated QTL position and the QTL-segregating families are given in Table 3.5.

3.4 Discussion

Growth and its manifestation, live weight, are quantitative traits of major importance for livestock species. Yet, although an easily recorded phenotype, live weight is a complex trait as it is the integral of all growth rates prior to that point, which in turn are a function of the animal's health, physiological state, strength of immune system and even its ability to compete for sometimes limited nutritional resources. Therefore, it would be beneficial to determine aspects of growth governed by genetic effects, and to utilize this information for genetic improvement. In order to dissect the genetic components of this trait: a) longitudinal live weight information

available for growing lambs was utilized and b) informative descriptors of growth and its components were defined.

Univariate QTL analyses of sheep live weight phenotypes at a particular point in time have been performed previously (e.g. Walling *et al.* 2004; McRae *et al.* 2005). However, isolation of live weights as single traits fails to capture the correlations between the components underlying growth. As a result, univariate studies have reduced power to detect QTL compared to techniques that combine information from multiple or longitudinal phenotypes.

Multivariate QTL models have previously been developed for longitudinal traits. These approaches have either transformed multivariate data into a single summary or composite measure (e.g. Weller *et al.* 1996; Gilbert & Le Roy 2003), or have modelled the time dependent QTL effects. As an example of the latter, Ma *et al.* (2002) and Wu & Hou (2006) described a maximum likelihood method for simple genetic structures (backcross, F2, full-sibs). In this method, the QTL was assumed to be a fixed effect with a specified number of alleles. This approach was applied to study growth QTL in mice and forest trees (Wu *et al.* 2004a,b; Wu *et al.* 2005). Alternatively, a longitudinal method using random regression (RR) was described by Lund *et al.* (2002) for animals and Macgregor *et al.* (2005) for human populations. Both modelled a QTL as a random effect using RR polynomials. With this approach, they analysed simulated data to assess QTL detection power of RR models fitted to longitudinal data with various time-dependence scenarios for the QTL. This method enabled the analysis of more general pedigrees. Recently, Lund *et al.* (2008) extended the RR model of the earlier simulation study, to allow for a genome scan and time-dependent QTL mapping for an actual trait. Their main objective was to

test the RR methodology against univariate QTL analysis of milk yield in dairy cattle.

Generally, the aforementioned "multivariate" approaches allowed for substantial increase in power for QTL detection in longitudinal data. Yet a number of issues arise when applying these techniques, mainly computational time and statistical difficulties in estimating many model parameters simultaneously. These issues are affected by the modelling choice for the QTL effect (fixed or random) and the need in random regression models to fit polynomials of high order for random QTL effects in order to capture fluctuations in the QTL variance over time. Additionally, it is usually difficult to assign biological meaning to the polynomials chosen to describe the QTL effect fitted in the model.

In the current chapter, an alternative approach was chosen for overcoming the limitations of univariate QTL studies of longitudinal traits. I sought to apply a multivariate method for the extraction of all phenotypic information present in longitudinal data and to then decompose the data in simpler but more informative variables for QTL analyses. For this purpose, live weight measurements over time were first modelled using growth curve functions, of which the Gompertz curve was found to be the most appropriate. By fitting the Gompertz model, information on weight measurements over time was used to estimate three parameters: A (weight at maturity); B (maximum growth rate); C (age at maximum growth rate). Since they are components of the same equation and estimated using the same data these parameters are correlated; thus, the use of the growth model allowed the reduction of the number of independent traits. Subsequently, predictions for weekly weights and growth rates were (non-linear and linear, respectively) combinations of the three

estimated model parameters. In this respect, a multiple testing issue was largely avoided when performing the QTL analyses.

Growth model equations had previously been employed to describe growth patterns, estimate growth parameters and dissect polygenic (and maternal genetic) components of growth variables in sheep (Lewis *et al.* 2002; Lewis & Brotherstone 2002; Lambe *et al.* 2006) and other species (Wang & Zuidhof 2004). In the current study, the application of growth models was extended by treating growth curve variables or predictions as traits for QTL studies. An analogous approach was employed by Rodriguez-Zas *et al.* (2002) and Minvielle *et al.* (2006) to detect QTL for dairy cattle milk traits and quail egg production traits, respectively, for parameters estimated from non-linear curves. For some of the QTL, Rodriguez-Zas *et al.* (2002) utilised the lactation equation to estimate trait values (e.g. milk yield, protein percentage) over time for alternative QTL allelic effects. Minvielle *et al.* (2006) identified significant QTL for different parameters of the egg laying curve. However, in contrast with the reported work, the aforementioned researchers did not make time-dependent predictions of the studied traits from the non-linear curves and, thus, trends in the overall QTL significance across time were not determined.

The procedure used in this chapter made no assumptions about the distribution of the QTL effects or their changes across time, and it allowed the detection of various QTL with different expression (significance and variance) patterns across time. QTL on chromosomes 3 and 14 had significant effects during early growth, whereas a QTL on chromosome 20 had significant effects on growth variables around intermediate/maximum growth. A growth rate QTL on chromosome 18 was only significant for later growth points (10 weeks). Further, the QTL on

chromosome 14 seemed to have contrasting allelic effects for early and late growth (around 24 weeks of age). Although, based solely on results from these analyses, the possibility that two distinct growth QTL were detected on chromosome 14 cannot be excluded, our approach would have the ability to detect QTL with alleles that have opposing action on growth in different stages of the animal's development. The phenotypes analysed in this dataset were a subset of those subjected to a heritability analysis by Riggio *et al*. (2008). The current results, in which QTL effects differ with age, are consistent with the results of Riggio *et al*. (2008) in which inter-age genetic correlations declined as the time period between weight measurements increased.

Another conclusion from the QTL analyses is that growth rate phenotypes may allow more effective detection of growth QTL effects than live weight phenotypes (either actual or predicted). This may well be the result of live weight being the more complex trait, i.e. the integral of all previous growth rates to that point in time. Furthermore, these analyses indicated that the same QTL have significant effects earlier for growth rate than for live weight, explicable by the fact that live weight is completely dependent on previous growth rates. This is apparent in the QTL significance trajectories for growth rate or predicted weight across age on chromosomes 14 and 20 (Figures 3.2 and 3.3). This shift to an earlier age for growth rate QTL on chromosomes 14 and 20 is observable because it lies within the age range of live weights present in our dataset and described by the growth curve. Failure to detect growth rate QTL on chromosome 3 and live weight QTL on chromosome 18 actually fits the observed pattern. A growth rate QTL on chromosome 3 would be expected to be significant prior to birth. In an analogous

manner, the live weight QTL on chromosome 18 would manifest itself at a later age point than the maximum age accurately covered by our data.

Overall, QTL mapping using growth curve descriptors seemed to be a reliable and efficient strategy. Growth QTL whose significance and effects varied with time were identified on distinct chromosomes at different growth stages. To confirm and enrich the results obtained, application of an alternative approach for QTL analysis on the same longitudinal phenotypes would be useful. Finally, implementation of the growth curve method to independent data would allow us to examine whether it can be generalised as a technique for mapping age-dependent growth QTL.

3.5 Conclusions

Fitting growth curves allowed the combination of information from multiple measurements into a few biologically meaningful variables, and the detection of growth QTL that were not observed from analyses of raw weight data. QTL analysis of growth parameters estimated from the Gompertz function provided important insight into growth as a multi-stage process in sheep. QTL significance varied with age, and distinct loci on different chromosomes seem to be active in at least three stages: early growth, intermediate/maximum and late growth. Additionally, this study revealed a trend by which loci associated with growth are apparent at a younger age for growth rate than for live weight. Finally, since distinct loci govern different growth stages, manipulation of the genetic factors underlying the different parts of an animal's growth curve to achieve distinct growth objectives may indeed be feasible.

Appendix 3.1

Genetic linkage map positions of genetic markers for the ovine chromosomes studied

Ovine chromosome 1

Marker Name	Position (cM)
BMS835	0
ILSTS44	29.2
MCM58	70.6
ILSTS29	87.1
BMS963	101.6
RM65	101.6
BM6438	117.4
BMS2321	142.4
MAF64	151.3
ILSTS004	158.4
CSSM04	191.6
BMS4000	201.0
INRA11	208.3
BMS527	214.1
DB6	217.7
BMS4001	226.1
BM7145	236.7
MCM137	237.3
BM6506	237.8
BMS4008	240.4
TGLA415	245.0
SOX2	247.1
BM8246	247.1
RM509	254.6
MCM130	260.9
BMS4045	270.4
BM864	270.4
CSSM32	272.1
LSCV105	275.9
BM1824	283.3
BMS1789	294.8
BM3205	310.1
OarHH36	410.1
URB014	410.1

Ovine Chromosome 2

Marker Name	Position (cM)
CSSM47	0
FCB226	47.6
BM3412	51.0
BMS1341	104
BL1080	121.4
BMS678	127.7
BMS1591	137.5
TGLA10	140.5
BM81124	192.9
CP79	216.1
TEXAN2	216.1
FCB20	245.4
BMS1126	263.2
BMS2626	305.7
BM6444	334.6
BM356	351.5
ARO28	365.7
FCB11	393.2

Ovine Chromosome 3

Marker Name	Position (cM)
BMS710	0
BMS2569	100
BM827	125.6
ILSTS42	130.6
ILSTS22	178.5
FCB5	182.4
BMC1009	192.6
KD0103	201.5
AGLA293	203.1
BL4	219.4
LYZ	219.4
IFNG	221.6
MAF23	234.2
CP43	241.5
CSRD111	251.5
TEXAN15	267.9
BM6433	279.9
BMS1248	282.7
BM8230	284.7
BMS772	299.2
BM2830	342.7

Ovine Chromosome 5

Marker Name	Position (cM)
TGLA176	0
RM006	11.2
TGLA48	20.3
TGLA303	25.8
BMS2258	78.0
BMS792	83.1
BM1853	94.0
SHP1	101.3
MCM527	109.2
CSRD2134	124.8

Ovine Chromosome 14

Marker Name	Position (cM)
TGLA357	0
TEXAN10	18.3
BMS2213	27.7
MT2	44.1
ILSTS10	58.1
BM8151	61.9
MCM133	63.9
INRA63	74.2
BM7109	79.4
BMS833	82.0
ILSTS002	90.8
LSCV30	117.5
MCMA19	136.3
BM6507	150.3

Ovine Chromosome 18

Marker Name	Position (cM)
ILSTS52	0
BP33	12.7
VH54	23.6
BMC5221	55.3
HH47	57.1
TGLA337	58.3
TGLA122	67.2
MCM38	71.4
ILSTS54	72.5
OB2	75.8
MCMA26	75.8
CSSM018	90.4
DLK	91.2
OY5	93.8
OY15	94.2

Ovine Chromosome 20

Marker Name	Position (cM)
INRA132	0
DYA	11.4
MCMA36	24.8
CP73	27.4
BM1815	39.7
OLADRB	54.0
OMHC1	56.5
BMS468	56.8
CSRD226	58.0
DRB1	59.1
TGLA387	65.1
BM1818	69.4
BP34	72.0
HH56	74.0
MCMA23	174

Ovine Chromosome 21

Marker Name	Position (cM)
BMC2228	0
ILSTS19	21.9
CP20	39.0
INRA175	39.0
VH110	42.8
JP15	48.5
BMC1206	81.4
BMS1948	86.7

CHAPTER 4

Further applications of the growth model approach using a simulated dataset of longitudinal phenotypes

4.1 Introduction

The growth curve technique developed in chapter 3 seems to be appropriate and efficient for analysing repeated measures for longitudinal traits in order to identify time-dependent QTL. Nevertheless, to verify the method's effectiveness and relevance for longitudinal phenotypes in general, it is necessary to evaluate it further by applying it to independent data, preferably using a dataset with known properties. Such an opportunity was presented by an independently simulated longitudinal dataset of yield, provided by the QTLMAS2009 workshop organisers for independent analysis by different groups. The structure of this dataset and the workshop objectives emphasised the need for development of quantitative genetic methods that would allow longitudinal phenotype analysis, for the purpose of identifying QTL with presumed time-dependence. Therefore, the application of the growth curve approach for QTL mapping to this dataset seemed appropriate.

As part of the growth model method, the Gompertz growth curve was again used to model the yield data, based on its good fit and predictive properties. Estimates of individual Gompertz curve parameters and predicted growth descriptors

(yields and growth rates) were employed in two separate collaborative studies, each with distinct objectives.

Firstly, a comprehensive set of analyses was done to detect QTL in the simulated population in order to compare the results of routinely used methods. Since the QTLMAS2009 data was structured in families, it allowed both linkage and association approaches to be evaluated, however this chapter focuses on linkage analyses using half-sib regression techniques. Specifically, sire and dam half-sib regression analyses were done a) on actual simulated yield at each time point, b) on individual curve parameters estimated after fitting the Gompertz curve to the longitudinal data, and c) on growth rates and yields predicted from the Gompertz model.

Secondly, a two-step approach was developed and applied, which combined growth curve fitting with genomic selection for longitudinal data. Genomic selection commonly refers to a new class of methods for genetic evaluation using dense marker maps that cover the entire genome (Meuwissen *et al.* 2001). The aim of this project was to estimate genomic breeding values for the trait at a time point (t600) which resided outside the range of longitudinal yield data provided. For this purpose, first the Gompertz function was fitted to the data for each individual. Then, a model for genomic selection was trained using the predicted phenotypes at t600 or the parameter estimates derived from the fitted Gompertz curve. Finally, genotype information for individuals without phenotypes (candidate set) was combined with the results of the genomic selection model to predict genomic breeding values at t600 for the candidate set.

Following the completion of the QTLMAS2009 workshop, retrospective comparisons of the detected QTL in the first study and the estimated genomic breeding values in the second study with the true simulated QTL and breeding values, respectively, allowed direct assessment of the growth curve approach for longitudinal data analysis, along with overall evaluation of its performance in comparison to that of other methods, employed independently by collaborators.

4.2 Materials and Methods

4.2.1 Data

The simulated data were provided by the QTLMAS2009 workshop organisers for analysis prior to the workshop's scheduled occurrence (20-21 April 2009) in Wageningen, The Netherlands. The dataset consisted of 2025 individuals from two generations. Of these, 25 individuals were parents, 20 female and 5 male. The remaining 2000 individuals were offspring from 100 full sibs (FS) families, with each FS family having 20 offspring. All individuals had complete marker information for 453 SNP marker loci which were randomly distributed over 5 chromosomes, each 1 Morgan in length. Offspring from 50 FS families (training set) had phenotype information of yield at 5 distinct time points (0, 132, 265, 397, and 530 days). The workshop organisers stated that yield values could represent weight during the growth of an animal or biomass during the growth of a crop. They also provided the range of asymptotic values of individuals' yield. The phenotyped FS families were chosen such that each female parent had at least 40 phenotyped

offspring while each male parent had 100 phenotyped offspring. The remaining offspring (candidate set) only had genotype information.

4.2.2 Gompertz growth model

I employed the Gompertz growth function to model the yield data over time, following the parameterization and procedure explained in chapter 3. Briefly, the Gompertz equation used is of the form: $y(t) = Ae^{\{-e[Be(C-t)/A]\}}$, where y(t) is the yield at time t; A the final yield; B the maximum growth rate and C the age at maximum growth rate.

The Gompertz model was first fitted across all trait data using nonlinear regression in SAS (release 9.1, SAS Inst., Cary, NC). It was then fitted to trait information for each individual separately using the nonlinear modelling procedure of JMP (release 7, SAS Inst., Cary, NC) to estimate individual model parameters A, B, C. Subsequently, the fitted individual equations and their derivatives were employed to predict yield and growth rate (yield per day), respectively, at the 5 time points for which trait information was available (0 to 530) and at time 600 (t600) which resided outside the given phenotypic range. In a specific part of the genomic evaluation study, genomic estimated breeding values (GEBVs) for the 3 model parameters were fitted for each individual in the Gompertz equation to predict GEBVs for yield at t600 from the growth function (further explained in the Genomic Evaluation section).

4.2.3 Exploratory data analyses

A colleague, Gib Hemani, tested the level of linkage disequilibrium (LD) among adjacent SNPs, using Haploview (Barrett et al. 2005). Specifically, the extent

of LD on each chromosome was determined by estimating the mean correlation, r, between alleles at adjacent SNP pairs and its square, r^2:

$$r_{ij} = D_{ij} / [p_i(1-p_i)p_j(1-p_j)]^{1/2} \text{ (Hill \& Robertson 1968),}$$

where $D_{ij}=p_{ij}-p_ip_j$ is the covariance of gametic frequencies, and p_i, p_j, p_{ij} are the frequencies of allele i at the first SNP locus, allele j at the second SNP locus and haplotype ij, respectively. Gib Hemani also performed univariate analyses to estimate heritabilities of yield at each time point and for the Gompertz model parameters using ASReml (Gilmour et al. 2002).

4.2.4 Half-sib QTL analyses

Half-sib QTL analyses were conducted using the data from the training set, pedigree information and parental genotypes, as described by Knott *et al.* (1996) and implemented in the web-based software GridQTL (Seaton *et al.* 2006). Analyses were performed separately for paternal and maternal half-sib families. Bruno Louro and Richard Leach performed dam and sire half-sib regressions, respectively, to map QTL for simulated yield at each time point. I also performed sire half-sib QTL analyses of the estimated Gompertz parameters A, B, and C and both sire and dam half-sib analyses of predicted yield and growth rates at the given time points and at t600. Bruno Louro conducted dam half-sib analyses of the model parameters A, B, and C. I subsequently repeated the analyses first conducted by Bruno Louro and Richard Leach to confirm the results and estimate relevant parameters.

In all analyses, empirical thresholds were obtained by permutation tests using 2000 permutations per chromosome. From these chromosome-wide thresholds the following significance levels were derived: chromosome-wide and genome-wide 5%

and 1%. For chromosomes where a single QTL had been identified a two-QTL model was evaluated and the best fitting two-QTL model obtained tested against the best one-QTL model.

Following identification of chromosome-wide significant QTL, the actual positions and variances of the simulated QTL were revealed by the QTLMAS organisers at the workshop. Thus, the performance of half-sib regression analyses of actual yields and predicted growth descriptors in identifying the actual QTL was subsequently assessed. For this purpose, the results from the half-sib analyses were scrutinised further, and nominally significant QTL ($P<0.05$; F-ratio>2.21 for sire and $F>1.57$ for dam analyses, respectively) were also identified, in addition to the ones reported previously (and initially submitted to the QTLMAS workshop).

4.2.5 Genomic evaluation

4.2.5.1 Procedure

To obtain genomic estimated breeding values (GEBVs) for yield at t600, Ricardo Pong-Wong combined genomic evaluation with the Gompertz growth model process that I followed for curve parameter estimation and trait prediction, using two different methods: I) estimating GEBVs for the Gompertz model parameters (A, B, C; i.e. 3 GEBVs per individual) and using them to estimate the breeding value for t600 from the Gompertz function; II) predicting the phenotypes at t600 by evaluating the Gompertz function with the estimated parameters and then applying genomic selection on the predicted t600 phenotypes (1 GEBV per individual). The training set was first used in the above procedure to evaluate the effect of each SNP genotype

and estimate GEBVs for each animal in this set. Subsequently, genomic evaluation was performed using the previously estimated SNP effects to obtain GEBVs separately for the candidate set and the parents, using both methods described above.

Following genomic evaluation, Pearson and Spearman rank correlations were calculated between GEBVs estimated from methods I and II, true and predicted phenotypes at each of the 5 given time points, and true and predicted yield at t600 for the training set, candidate animals, and parents, as well as all animals as a whole. The Pearson correlation coefficient is the standard parametric correlation, and the Spearman's rank correlation is simply the correlation of the ranked variables.

4.2.5.2 Genomic evaluation model

A Bayes B type of analysis was used as first described by Meuwissen *et al.* (2001). Under a Bayesian framework the model accounts for the fact that not all SNPs affect the trait in question. The model assumed in the method is:

$$\mathbf{y} = \mathbf{Xb} + \sum_{i=1}^{m} \mathbf{g}_i \alpha_i + \mathbf{e}$$

where **y** is the vector of phenotypes; **b** contains the fixed effects and **X** is its incidence matrix; α_i is the allelic substitution effect for SNP *i*; $\mathbf{g}_i$ is the vector of genotypes (1, 2 & 3 for genotypes 00, 10/01 and 11, respectively) for SNP *i*; and **e** the vector of residuals distributed $N(0, \sigma_e^2)$. The allelic substitution effects α for each SNP are assumed to be from a mixture distribution with probability π of having an effect on the trait, and hence a probability $(1-\pi)$ of not affecting the trait. If the SNP is affecting the trait, its allelic substitution is distributed $N(0, \sigma_{snp}^2)$, with σ_{snp}^2 being common across SNPs with an effect.

The implementation of the model was done using Gibbs sampling (Wang *et al.* 1993,1994). Assuming flat priors for σ_e^2, σ_{snp}^2 and π, and given the likelihood function of the data, the conditional distributions are scaled inverted chi-square for the two variances and a beta distribution for π (Wang *et al.* 1993,1994; Janss *et al.* 1995). So far, the implementations of Bayes B reported in the literature have not estimated π, but assumed it was known without error (e.g. Meuwissen *et al.* 2001; de Roos *et al.* 2007). In this model, π was estimated directly from the data. This constitutes further development of the Bayes B approach for genomic evaluation.

For each analysis, a MCMC chain was run and the first 10000 cycles were discarded as burn-in period. Following this, 10000 realisations were collected, each separated by 50 cycles between consecutive realisations. The posterior mean was used as the estimate for each parameter of interest.

4.3 Results

4.3.1 Exploratory data analyses

The LD between adjacent SNP pairs was generally low. Chromosomes 1 to 4 had similar distributions of square correlation values (r^2), with the estimated mean r^2 between adjacent markers being ~0.15, but chromosome 5 appeared to have much lower LD (Appendix 4.1).

The estimated yield heritability was ~0.50 for all time points, varying between 0.46 and 0.50. The heritabilities for the curve parameters A, B and C were 0.45, 0.48 and 0.26, respectively.

4.3.2 Gompertz growth model parameters

The Gompertz model provided a good fit of the data as a whole (Figure 4.1 and Appendix 4.2) and when the curve was fitted for each individual. To further test how well the Gompertz curve fitted the phenotypic data, yield phenotypic values were predicted at all 5 time points for which observed phenotypic data was available. The Pearson and Spearman correlations between observed and predicted phenotypic values (calculated by Ricardo Pong-Wong) were above 0.99 at time 530 (t530), with similar high correlations obtained for the other 4 time points. These high correlations remained when comparing the GEBVs calculated for both the true and predicted phenotypes, with correlations ≥0.98 at all five time points. Overall, these results suggested that the data had likely been simulated with a parametric curve related to the Gompertz.

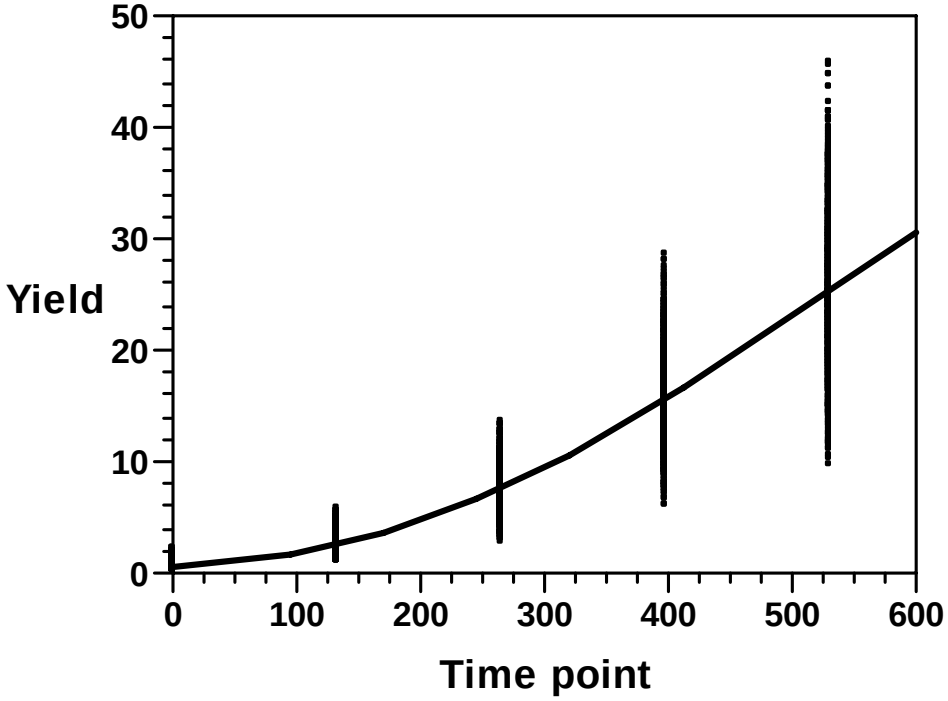


Figure 4.1 Distribution of yield of individuals across time (vertical bars/dots) and the average growth curve obtained after fitting the Gompertz model to the yield data (solid line).

4.3.3 Half-sib QTL analyses

The results of sire and dam half-sib analyses of univariate yield and Gompertz growth descriptors are summarised in Tables 4.1-4.3. At a minimum, chromosome-wide significance ($P<0.05$) was used as a criterion to report QTL in the above tables, since at the first part of the analysis, no knowledge was available about the actual simulated QTL in terms of number of QTL on each chromosome, QTL position and variance. F-statistic curves over time for each QTL are presented for all half-sib analyses in Hadjipavlou *et al.* (2010). Chromosome- and genome-wide significant QTL were identified for all time points and all chromosomes, with the QTL significance varying across time for most of the detected loci. Analyses of the Gompertz descriptors (i.e. curve parameters, predicted growth rates and yields) identified all QTL detected by half-sib regression on univariate yields (Tables 4.1-4.3), apart from a significant QTL at 38cM on chromosome 4 (Table 4.1). Additionally, Gompertz descriptor analyses mapped a QTL at 99cM on chromosome 3 (Table 4.2), not found by analyses of actual yield.

Overall, both actual yield and growth descriptor analyses mapped one QTL on chromosome 1, three QTL on chromosome 2, and two on each of chromosomes 3, 4 and 5 (Tables 4.1-4.3). On the whole though, the growth model method seemed to provide more information about the nature of each QTL. For example, some QTL were found to be significant for a curve parameter (i.e. affected growth curve shape). Other QTL, such as one on each of chromosomes 2 and 4, had different time-range significance for accumulated yield or instantaneous growth rate. These QTL were significant at earlier ages for predicted growth rate than for predicted yield.

4.3.4 Description of the true model used to simulate data and QTL effects

Subsequent to completing the analyses, the model used to simulate the data was revealed as being the logistic growth curve, (Coster *et al.* 2010) described as $y(t) = \phi_1 / \left[1 - e^{(\phi_2 - t)/\phi_3}\right]$, where y(t) is the yield over time, ϕ_1 the asymptotic trait value, ϕ_2 the time at the inflection point of the curve (time at maximum growth rate), and ϕ_3 the relative growth rate. A total of 18 QTL were simulated directly on the logistic curve parameters: one distinct QTL per parameter on each of the 5 chromosomes, and a second QTL for ϕ_1 on chromosome 2, for ϕ_2 on chromosome 3, and ϕ_3 on chromosome 4 (Tables 4.4-4.6). All QTL had purely additive effects on the trait, i.e. no dominance, imprinting or epistatic effects were simulated.

Table 4.1 Summary of QTL results from univariate half-sib analyses of observed yield

Chromosome	QTL Position (cM)[1]	F-ratio[2]	Time Range of Significance[2]	% Vp Explained by QTL[3]
Paternal half-sib analyses				
1	41-43	8.01-10.87	0-530	21
2	5-6	3.31-6.2	0-530	12
2	91	4.24-4.69	0-530	9
3	27	4.65	0	9
4	78-79	4.13-7.57	0-397	13
5	76	4.22-4.52	397-530	8
Maternal half-sib analyses				
1	44-45	6.01-7.48	0-530	47
2	39	2.13-2.99	0-530	16
3	78-79	2.32-2.44	0-530	11
4	5-8	2.34-3.08	0-132	18
4	38	2.01	265	8
4	86	2.08-2.57	0-132	13
5	99	1.97-2.25	0-265	10

[1]QTL position is defined relative to the first marker (SNP) in the genetic map for each chromosome; first marker positioned at 1cM.

[2]F-statistics given as range across time points for which the QTL was deemed significant chromosome-wise with at least P-values < 0.05 (determined by permutation testing).

[3]Vp refers to the phenotypic variance for each trait. The proportion of Vp due to the QTL was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$, where RMSQ is the residual mean square estimate for each model and "full" is the model with the QTL effect fitted and "reduced" is the model without the QTL effect (Knott *et al.* 1996). For each QTL, it was estimated at the time point with the highest test statistic.

Table 4.2 Summary of QTL results from paternal half-sib analyses of growth descriptors, predicted from the Gompertz model

Gompertz Growth Descriptor	Chr.	QTL Position (cM)[1]	F-ratio[2]	Time Range of Significance[2]	% Vp Explained by QTL[3]
Yield	1	39-42	8.16-11.8	0-600	21
Growth rate	1	40-44	7.75-17.13	0-600	30
Model Parameters	1	38-40	14.64-22.14	A, B, C	38, 28, 26
Yield	2	6	7.84-7.92	530-600	13
Yield	2	49	7.12-7.94	265-397	13
Yield	2	91-97	3.97-5.35	0-132	9
Growth rate	2	2-6	7.88-7.95	397-600	13
Growth rate	2	49	7.37-8.47	132-265	15
Growth rate	2	91	4.83	0	4
Model Parameters	2	1-6	7.53-8.0	A, B, C	13, 14, 5
Model Parameters	2	35	3.14	C	4
Yield	3	27-28	4.23-5.96	0-132	10
Growth rate	3	27	4.97	0	4
Growth rate	3	99	3.00	265	4
Yield	4	79	3.18-7.23	0-600	12
Growth rate	4	79	3.27-7.27	0-265	10
Yield	5	76	3.67-4.59	265-600	7
Growth rate	5	75-76	3.66-4.86	132-600	8
Model Parameters	5	74-75	4.09-4.87	A,B	6, 8

[1]QTL position is defined relative to the first marker (SNP) in the genetic map for each chromosome; first marker positioned at 1cM.

[2]F-statistics given as range across time points for which the QTL was deemed significant chromosome-wise with at least P-values < 0.05 (determined by permutation testing).

[3]Vp refers to the phenotypic variance for each trait. The proportion of Vp due to the QTL was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$, where RMSQ is the residual mean square estimate for each model and "full" is the model with the QTL effect fitted and "reduced" is the model without the QTL effect (Knott *et al.* 1996). For each QTL, it was estimated at the time point with the highest test statistic.

Table 4.3 Summary of QTL results from maternal half-sib analyses of growth descriptors, predicted from the Gompertz model

Gompertz Growth Descriptor	Chr.	QTL Position (cM)[1]	F-ratio[2]	Time Range of Significance[2]	% Vp Explained by QTL[3]
Yield	1	43-44	4.99-7.46	0-600	47
Growth rate	1	43-45	5.89-7.38	0-600	46
Model Parameters	1	43-44	6.34-7.15	A, B	39, 45
Yield	2	36	1.94-2.6	0-530	13
Growth rate	2	2	1.98-2.05	530-600	8
Growth rate	2	37	2.16	0	11
Model Parameters	2	2	1.99-2.09	A, B	8, 8
Yield	3	79	2.08-2.16	265-397	9
Yield	3	87	2.06	530-600	8
Yield	3	95	2.16-2.19	0-132	10
Growth rate	3	78-79	2.08-2.31	0-397	11
Growth rate	3	87	2.02-2.04	530-600	8
Model Parameters	3	78	2.06	A	8
Model Parameters	3	87	2.04	B	8
Yield	4	7	2.53-3.35	0-132	18
Growth rate	4	3	2.59	0	11
Model Parameters	4	82	1.91	C	7
Yield	5	99	1.93-2.33	0-265	11
Growth rate	5	99	2.24	0	11

[1]QTL position is defined relative to the first marker (SNP) in the genetic map for each chromosome; first marker positioned at 1cM.

[2]F-statistics given as range across time points for which the QTL was deemed significant chromosome-wise with at least P-values < 0.05 (determined by permutation testing).

[3]Vp is the phenotypic variance for each trait. The proportion of Vp due to QTL was estimated as $4x(1-RMSQ_{full}/RMSQ_{reduced})$, where RMSQ is the residual mean square estimate for each model. “Full” is the model with the QTL effect fitted and “reduced” the model without the QTL (Knott *et al.* 1996). For each QTL, it was estimated at the time point with the highest test statistic.

4.3.5 Comparison of detected QTL with actual simulated QTL effects

Tables 4.4 and 4.5 summarise retrospective comparisons of the performance of the half-sib methods in identifying the actual simulated QTL effects. Upon acquiring knowledge on the actual QTL for each chromosome, both sire and dam half-sib regression results of actual yields and growth descriptors were re-examined by determining the nominal threshold for QTL significance using the F-ratio for a single test (P-value<0.05). Additional QTL of nominal significance were found and these are reported in Tables 4.4-4.5 in addition to the (at least) chromosome-wide significant QTL given in Tables 4.1-4.3, to allow comprehensive comparisons.

Half-sib analyses of both actual yield and predicted growth descriptors identified a nominally significant QTL on chromosome 5 (within 10cM of actual position at 60cM; Tables 4.4-4.5). Another QTL was identified at 88cM (±10cM) on chromosome 1 only from the sire analysis of actual yield. The Gompertz model descriptors identified four additional QTL of nominal significance: one for parameter C on chromosome 1 (actual position at 54cM; Table 4.5), another one for predicted growth rate on chromosome 3 (actual QTL located at 56cM; Table 4.4), and one on each of chromosomes 4 and 5 (Table 4.5; QTL positioned at 70 and 31cM, respectively). The latter were detected for asymptotic yield (Gompertz parameter A) and maximum growth rate (Gompertz parameter B), respectively. None of these four QTL was found with univariate analysis of yield (Table 4.4).

In total, out of 18 simulated QTL, half-sib analyses of yield at each time point allowed detection of 10 QTL with chromosome-wide significance or higher (Tables 4.1, 4.4), two nominally significant ones (Table 4.4) and one false positive QTL

(chromosome 5 at 99cM; Table 4.1). Analysis of growth model descriptors (yield, growth rate and the Gompertz model parameters) resulted in 10 QTL all of which had at least chromosome-wide significance, and six QTL of nominal significance. The same false positive QTL identified by the actual yield analyses was detected (chromosome 5 at 99cM; Tables 4.2-4.3). Finally, half-sib regression analyses of both yield and growth descriptors resulted in variance overestimation for some QTL (Tables 4.4-4.5).

Table 4.4 Correspondence between simulated QTL positions and variances with those of QTL effects mapped from half-sib analyses of actual yield and predicted growth rate

Chr.	Param.[2]	Actual QTL Position (cM)	Actual QTL Var. (%Vp)	Sire half-sib regression analysis[1]		Dam half-sib regression analysis[1]	
				Yield	Growth Rate	Yield	Growth Rate
1	Φ_1	42	29.3	**21	**30	**47	**46
1	Φ_2	54	32.3				
1	Φ_3	88	23.4	♦˙5			
2	Φ_1	5	4.6	**12	**13		*8
2	Φ_2	33	33.0			**16	*11
2	Φ_3	49	48.9		**15		
2	Φ_1	89	3.7	**9	**4	♦˙5	
3	Φ_2	7	3.5				
3	Φ_3	26	4.7	**9	**4		
3	Φ_2	56	3.8		♦˙3		
3	Φ_1	90	4.1	♦˙5	*4	**˙11	*8
4	Φ_3	10	5.9			**18	**11
4	Φ_2	37	3.2			*8	
4	Φ_1	70	3.3				
4	Φ_3	86	6.6	**˙13	**10	**13	
5	Φ_3	31	4.6				
5	Φ_2	60	3.7	♦˙5			
5	Φ_1	77	2.5	**8	**8		♦˙6

[1]QTL with: **genome-wide; *chromosome-wide; ♦nominal significance. ˙QTL position was estimated more than 5cM (but not more than 10cM) away from the actual position. The estimated % Vp explained by each QTL for actual yield or predicted growth rate is given.

[2]Parameters from logistic function used to simulate data: asymptote (Φ_1), time at inflection point (Φ_2) and relative growth rate (Φ_3).

Table 4.5 Correspondence between simulated QTL positions and variances with those of QTL effects mapped from half-sib analyses of Gompertz model parameters

Chr.	Param.[2]	Actual QTL Position (cM)	Actual QTL Var. (%Vp)	Sire Half-sib Regression Analysis[1] Gompertz Parameters			Dam Half-sib Regression Analysis[1] Gompertz Parameters		
				A	B	C	A	B	C
1	Φ_1	42	29.3	**38	**28	**26	**39	**45	
1	Φ_2	54	32.3			♦·3			
1	Φ_3	88	23.4						
2	Φ_1	5	4.6	**13	**14	*5	* 8	* 8	
2	Φ_2	33	33.0			*4	♦·6	♦·6	
2	Φ_3	49	48.9						
2	Φ_1	89	3.7	♦·4	*5				
3	Φ_2	7	3.5						
3	Φ_3	26	4.7	♦·3		♦·3			
3	Φ_2	56	3.8		♦·3				
3	Φ_1	90	4.1				*·8	*8	
4	Φ_3	10	5.9						
4	Φ_2	37	3.2						
4	Φ_1	70	3.3	♦·4					
4	Φ_3	86	6.6						*7
5	Φ_3	31	4.6					♦·5	
5	Φ_2	60	3.7	♦·3	♦·4				
5	Φ_1	77	2.5	**6	**8				

[1]QTL with: **genome-wide; *chromosome-wide; ♦nominal significance. ·QTL position was estimated more than 5cM (but not more than 10cM) away from the actual position. The estimated % of Vp explained by each QTL for each Gompertz parameter is given.

[2]Parameters from logistic function used to simulate data: asymptote (Φ_1), time at inflection point (Φ_2) and relative growth rate (Φ_3).

4.3.6 Estimation of GEBVs for Gompertz curve parameters

Univariate genomic evaluation analyses were performed on each of the 3 parameters of the Gompertz function. The correlations between the univariate GEBVs for the parameters were high; correlations between GEBVs for A-B, A-C and B-C were 0.97, 0.71 and 0.59, respectively. The posterior means of π for A, B and C were 0.059, 0.082 and 0.219, respectively. The posterior probabilities for the

SNPs having an effect on parameters A, B and C, and their estimated allelic substitution effects are shown in Figures 4.3 and 4.4, respectively. The results suggest that parameters A and B are largely affected by the same SNPs, with some others affecting parameter C. This is consistent with the high correlation between the GEBVs for A and B.

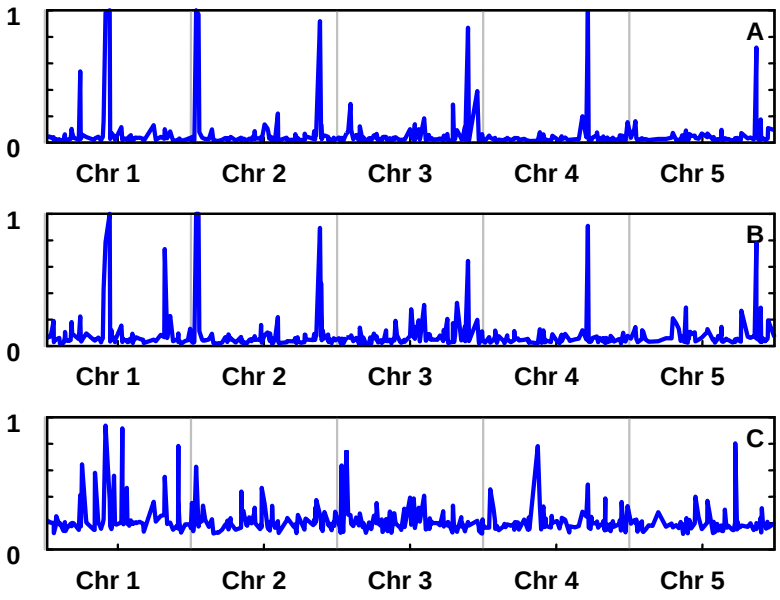


Figure 4.3 Probability of an individual SNP affecting parameters A, B or C.

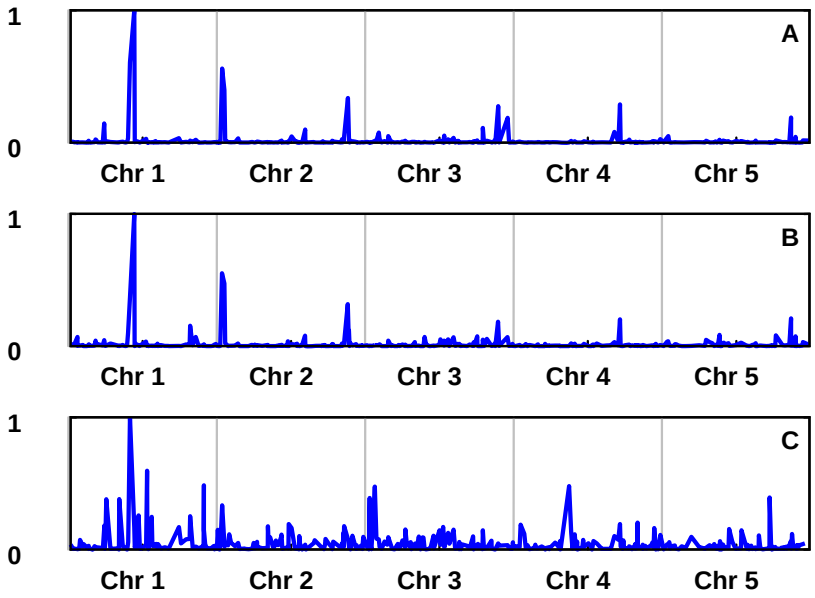


Figure 4.4 Estimated allele substitution effect for A, B and C. The size of the effects are rescaled relative to the largest allele effect within each parameter (i.e. highest effect=1).

4.3.7 Estimation of GEBVs for yield at a given time point

The GEBVs for t600 obtained by evaluating the Gompertz function with GEBVs for A, B and C (method I) were very similar to those calculated from method II in which breeding values were directly estimated for predicted performance at t600 (Figure 4.5). The correlation between both approaches for calculating GEBVs was 0.99, with GEBVs from method I having slightly larger variance. The GEBVs obtained from genomic selection on the predicted trait at t600 show a very similar trend as found for parameters A and B, with the same SNPs of large effect found for A and B also affecting t600 (Appendix 4.3). The estimate of π for t600 was 0.048.

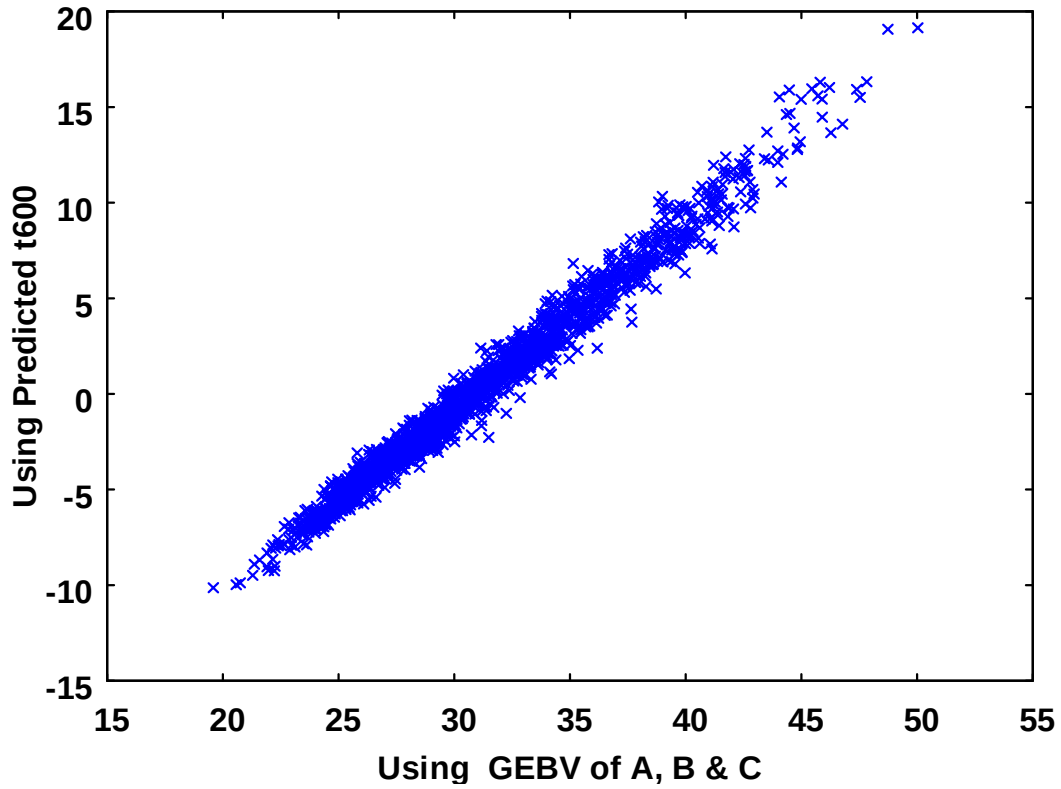


Figure 4.5 Scatter plot of t600 GEBVs calculated by evaluating the Gompertz function using GEBVs of A, B, C (x-axis) and from genomic selection on the predicted phenotype at t600 (y-axis). The GEBVs not scaled on the same mean.

4.3.8 Comparison of the fitted with the true model used for the simulation

Even though the Gompertz model fitted the data well, the logistic growth curve was actually used to simulate the trait, and QTL were mapped directly on the logistic curve parameters. Both the logistic and the Gompertz models are characterised by an 'S' shape growth with an asymptotic maximum yield, but the parameters describing them have different meaning. Comparing both functions, parameters A and ϕ_1 have a similar definition. The equivalence between the other

parameters is less clear, which explains the results obtained. GEBVs for A, B were highly correlated to the true values for ϕ_1, but poorly with ϕ_2 and ϕ_3 (Table 4.6.; for correlations within training and candidate sets see Appendix 4.4).

Analysis of Gompertz parameters A and B showed that the SNPs with the highest probability of having an effect on the trait were located at the positions where the six QTL affecting ϕ_1 were simulated (Figures 4.3-4.4). Locations containing QTL for parameters ϕ_2 and ϕ_3 were less associated with QTL affecting the parameters of the Gompertz curve.

Table 4.6 Pearson (lower diagonal) and Spearman (upper diagonal) correlations between true and estimated breeding values for t600 and the parameters used to simulate or analyse the data for all animals

	TBV[1]				**GEBV[2]**				
	T600	**Φ_1**	**Φ_2**	**Φ_3**	**t600_I**	**t600_II**	**A**	**B**	**C**
TBV t600		0.995	0.230	0.091	0.935	0.937	0.913	0.930	0.405
TBV Φ_1	0.997		0.285	0.160	0.931	0.937	0.928	0.925	0.465
TBV Φ_2	0.291	0.344		0.129	0.237	0.258	0.377	0.306	0.719
TBV Φ_3	0.098	0.157	0.108		0.082	0.112	0.213	0.029	0.463
GEBV t600_I	0.942	0.941	0.316	0.079		0.990	0.968	0.979	0.402
GEBV t600_II	0.947	0.949	0.332	0.116	0.990		0.969	0.981	0.437
GEBV A	0.919	0.933	0.459	0.194	0.970	0.969		0.957	0.599
GEBV B	0.938	0.940	0.396	0.034	0.983	0.983	0.971		0.454
GEBV C	0.519	0.571	0.735	0.433	0.523	0.551	0.709	0.587	

[1]TBV: true breeding values for yield at time 600 (t600) and logistic growth curve parameters Φ_1, Φ_2, Φ_3.

[2]GEBV: genomic breeding values for t600 estimated with method I and II (t600_I, t600_II, respectively) and for Gompertz curve parameters A, B and C.

4.3.9 Comparison of estimated and true GEBVs at time 600

Although the Gompertz curve was not the true model, its use provided very accurate GEBVs for t600. The correlations between the true breeding values and GEBVs for all animals are presented in Table 4.6 and for the training, candidate sets and parents, separately, in Appendix 4.4. Both methods of estimating GEBVs yielded similar accuracy. The Pearson and Spearman correlations between true and estimated breeding values from both methods I and II were above 0.93 for all individuals in the pedigree (Table 4.6).

4.4 Discussion

4.4.1 Application of a growth model approach to identify QTL and perform genomic evaluation for a longitudinal trait

Analysis of an independently simulated dataset resulted in further validation and provided evidence for the usefulness of the growth curve approach for QTL mapping of longitudinal data. QTL mapping using half-sib analyses of individual Gompertz curve parameter estimates, predicted yields and growth rates succeeded in capturing 16 out of the 18 simulated QTL with one false positive. Overall, detecting QTL for growth descriptors obtained from the Gompertz model provided good description and strong statistical support for growth QTL whose significance and variance changed with time. This was true despite the fact that the logistic growth curve had been used to simulate the QTL.

Further, a novel application of the growth model approach was unveiled, as part of a two-step procedure aiming at genomic evaluation for a longitudinal trait. In this approach, Gompertz growth curve modelling of the simulated longitudinal dataset was performed first and then genomic evaluation was applied on the derived model parameters in one method and on a trait value predicted from the growth model, in another method. High correlations between true and predicted genomic breeding values showed that implementing the growth modelling step separately from genomic evaluation provided a simple way for genomic evaluation on a longitudinal trait without any detrimental effect on breeding value estimation. In addition, the method for GEBV estimation by evaluating the Gompertz function using GEBVs of the three model parameters could be of great benefit for cases when GEBVs are needed for different time points.

4.4.2 Performance of the growth model approach versus univariate half-sib analysis of actual data in mapping the actual simulated QTL effects

Because the Gompertz growth curve fitted the yields well, half-sib regression analysis of yields and growth rates predicted from the Gompertz model performed better in identifying the simulated QTL effects than linkage analysis of univariate phenotypes. In addition to detecting more of the simulated QTL, analysis of the Gompertz predictors provided an insight into the nature of the QTL, with some being significant solely for curve parameters, and others being detectable for predicted growth rate and/or yield at particular time ranges.

Even though half-sib analyses of the growth model descriptors mapped loci with higher concordance with the simulated QTL, it is worth noting that the advantage of using the Gompertz model approach was more pronounced upon examination of the cumulative QTL results of all growth predictors (curve parameters, predicted yields and growth rates). QTL that were detected for each of the Gompertz parameters did not clearly correspond to QTL simulated for each of the logistic model parameters. This would be expected since a) only Gompertz parameter A had similar definition to parameter Φ_1, describing asymptotic yield in the logistic model; b) each of these two parameters had different correlations with the other two in the respective model, based on the distinct properties of the two growth models.

4.4.3 Performance of other methods in mapping the actual simulated QTL effects

In addition to the half-sib regression analyses, univariate variance component (VC) and SNP association analyses were done on yield to map the simulated QTL. Suzanne Rowe performed the VC analyses, whereas Gib Hemani and Javad Nadaf conducted the association study. The association and VC methods detected 6 and 10 out of the 18 simulated QTL, respectively, with 2 and 5 false positives. Moreover, the association analysis tended to underestimate the phenotypic variance attributed to the QTL with the VC analysis giving the most accurate estimation of variance explained (Appendix 4.5). For completeness, the number of QTL out of the 18 simulated QTL detected by each approach are summarised in Appendix 4.6. However, it is important to note that the VC and association methods were not

applied on the Gompertz curve predictors. Therefore, a full comparison of their performance with that of the half-sib analysis of the growth model parameters and predictions is not applicable.

4.4.4 A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data

Gompertz curve fitting and predictions made it possible to estimate genomic breeding values for a time point with no phenotypic records, time 600, both for individuals with trait information at other time points (training set) and individuals with no phenotypes (candidate set). For the training set, genomic breeding values calculated from predicted phenotypes were highly correlated with breeding values obtained by directly using the respective observed phenotypes. Retrospective comparison revealed high accuracies between true and estimated genomic breeding values at time 600; these were above 0.93 for both the training and candidate sets.

In this study, the proportion of SNPs affecting the trait, parameter π, was estimated in the process of genomic evaluation. This contrasts with previously reported implementations of Bayes B where π was assumed to be known without error. The π values for parameters A, B, C and yield at t600 were overestimated, possibly partly due to the low LD between SNPs (average r^2 between consecutive SNPs was 0.15). With insufficient marker density, low LD would imply that more than one SNP adjacent to the QTL would need to be assigned an effect in order to explain the total QTL variance. However, it is necessary to bear in mind that the Gompertz function was not the true model for the data, and this may have further

complicated the interpretation of the estimated π values. Nonetheless, the success in estimating such an important parameter from the data itself provides an improvement in genomic evaluation relative to assuming that π is known without error.

The correlation between GEBVs for t600 obtained using method I with GEBVs estimated with method II was very high (0.99) for both training and candidate sets. This high correlation is in agreement with the finding that the same SNPs of large effect found for Gompertz parameters A and B also affected performance at t600. Based on these findings, method I may be preferable in cases that GEBVs for growth are required at different time points. With method I, the genomic evaluation step only needs to be performed three times for each animal; each time individual GEBVs will be obtained for one of the model parameters. Subsequently, the 3 GEBVs per animal can be used to evaluate the Gompertz growth function and obtain GEBVs at any time point of interest.

However, investigation and assessment of both methods is necessary, so as to confirm that the same major QTL affect the model parameters and the trait at the time point(s) of interest. If this is not pursued, it is possible to ignore QTL with significant effects that might not affect the model parameters, especially in the case that a suboptimal growth model is chosen to fit the longitudinal phenotypes. In this study, even though the logistic, and not the Gompertz model, was actually used to simulate the dataset, the time point of interest (t600) resided close to the point of maximum growth rate, a region with highest coincidence between the true and fitted models. The outcomes might have been very different if predictions were made at regions where the suboptimal Gompertz model did not fit the data adequately.

4.5 Conclusions

A growth curve approach for age-dependent QTL mapping was further validated in a collaborative project, in which a combination of intra- and inter-family based methods that exploited linkage or linkage disequilibrium for QTL mapping were used on simulated longitudinal data. Retrospective assessment of the ability of different methods to map the simulated QTL showed that QTL identification using the Gompertz curve predictors performed better than the other approaches, as it succeeded in detecting most of the simulated QTL, with only one false positive, which had also been mapped using univariate analysis of actual yield. This outcome was observed despite the fact that the data were actually simulated using a logistic function. Fitting the data using a growth curve resulted in the correlation among measurements ordered in time being taken into account for parameter estimation and trait prediction. This, in turn, led to more informative phenotypes for QTL analyses than the univariate measurements used in other approaches. Accounting for the longitudinal aspect of the trait proved important for QTL detection since the significance and variance changed with time for the majority of simulated QTL.

Extension of the growth model method resulted in a two-step approach that combined curve fitting and genomic selection on longitudinal data. This procedure proved to be a simple and reliable strategy. Despite that the Gompertz curve was not the true model used to simulate the data, the correlations between true and estimated breeding values for yield at time 600 were high (Pearson and Spearman correlations above 0.93) for both methods employed. The method of using genomic estimated breeding values (GEBVs) of the three model parameters to evaluate the Gompertz

function in order to estimate GEBVs for the trait at a time of interest could be beneficial when GEBVs are needed for different time points. In this study, the proportion of SNPs affecting the trait was estimated from the data, contrasting with previous implementations of genomic selection in which this proportion had been assumed to be known without error. The findings of this study showed that separate performance of the growth modelling and genomic evaluation processes led to great simplification of the methodology with no detrimental effect on the final results.

Appendix 4.1

Analysis of LD structure for each chromosome

Chromosome	Mean r^2 between adjacent markers	Mean maximum pair-wise r^2 for each marker
1	0.146	0.506
2	0.154	0.582
3	0.159	0.523
4	0.124	0.496
5	0.082	0.387

Appendix 4.2

Estimated means and standard errors for Gompertz model parameters and predicted weight at point 600[1].

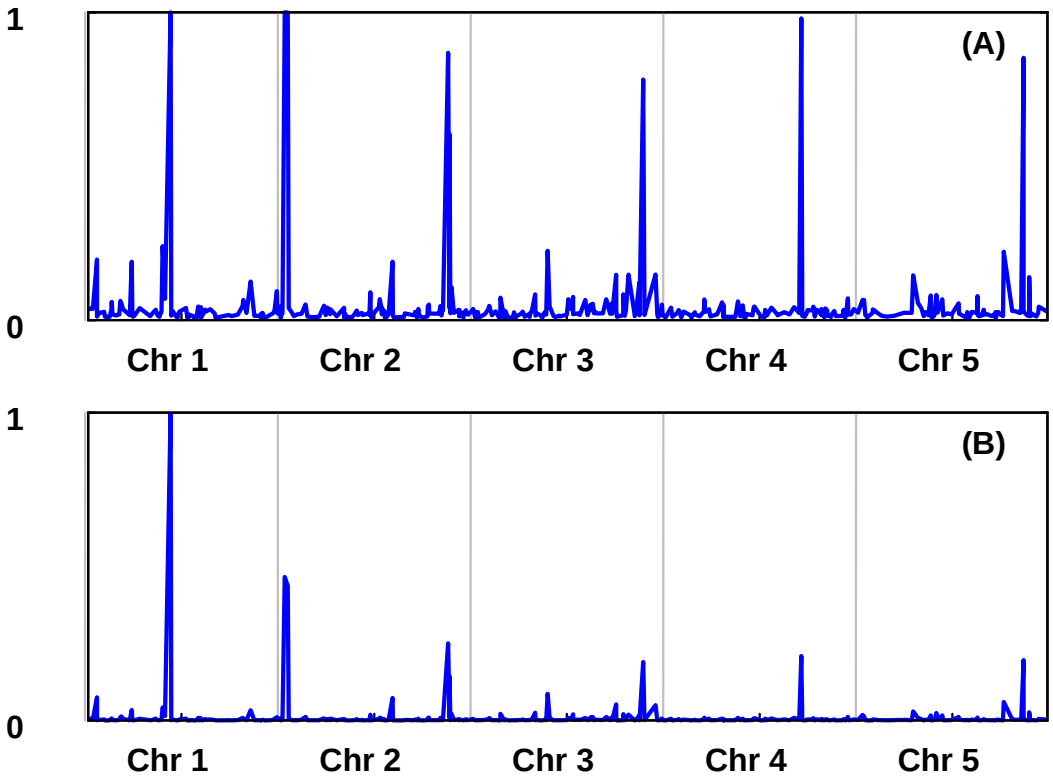
Parameter	Sample mean	Approx. SEM[2]	Approx. 95% Confidence Interval
A (yield)	67.9	0.65	66.58-69.13
B (yield/time unit)	0.076	0.001	0.075-0.077
C (time)	520	1.23	517.8-522.6
Yield at time 600	30.6	0.24	30.11-31.05

[1]The average of individual means, predicted after fitting the Gompertz model to trait values for each individual, was estimated for each parameter.

[2]SEM=Standard Error of the Mean.

Appendix 4.3

(A) Probability of a SNP having an effect (π) and (B) allele substitution effect (α) on the performance at time 600.



Appendix 4.4

Pearson (lower diagonal) and Spearman (upper diagonal) correlations between true and estimated breeding values for t600 and the parameters used to simulate or analyse the data[1,2].

Unphenotyped offspring (candidate set)									
	TBV[1]				**GEBV[2]**				
	t600	**Φ_1**	**Φ_2**	**Φ_3**	**t600_I**	**t600_II**	**A**	**B**	**C**
TBV t600		0.995	0.203	0.036	0.930	0.934	0.899	0.923	0.364
TBV Φ_1	0.997		0.259	0.101	0.927	0.934	0.914	0.919	0.422
TBV Φ_2	0.292	0.349		0.118	0.235	0.246	0.376	0.284	0.737
TBV Φ_3	0.051	0.106	0.093		0.031	0.057	0.153	-0.031	0.389
GEBV t600_I	0.940	0.941	0.339	0.040		0.989	0.967	0.979	0.387
GEBV t600_II	0.945	0.948	0.350	0.067	0.990		0.965	0.980	0.415
GEBV A	0.912	0.928	0.485	0.136	0.971	0.967		0.952	0.584
GEBV B	0.935	0.938	0.406	-0.014	0.984	0.984	0.970		0.424
GEBV C	0.503	0.555	0.758	0.353	0.533	0.551	0.713	0.586	
Phenotyped offspring (training set)									
	TBV				**GEBV**				
	t600	**Φ_1**	**Φ_2**	**Φ_3**	**t600_I**	**t600_II**	**A**	**B**	**C**
TBV t600		0.995	0.258	0.139	0.940	0.940	0.926	0.935	0.443
TBV Φ_1	0.997		0.311	0.210	0.935	0.939	0.939	0.931	0.503
TBV Φ_2	0.294	0.343		0.138	0.243	0.274	0.378	0.330	0.696
TBV Φ_3	0.137	0.198	0.132		0.127	0.162	0.268	0.083	0.530
GEBV t600_I	0.944	0.941	0.294	0.113		0.991	0.968	0.979	0.417
GEBV t600_II	0.949	0.950	0.317	0.157	0.990		0.972	0.981	0.460
GEBV A	0.925	0.939	0.435	0.244	0.969	0.970		0.962	0.614
GEBV B	0.941	0.941	0.389	0.076	0.983	0.983	0.972		0.482
GEBV C	0.534	0.586	0.717	0.500	0.517	0.554	0.707	0.591	
Parents									
	TBV				**GEBV**				
	t600	**Φ_1**	**Φ_2**	**Φ_3**	**t600_I**	**t600_II**	**A**	**B**	**C**
TBV t600		0.985	0.099	0.092	0.916	0.910	0.885	0.914	0.315
TBV Φ_1	0.997		0.179	0.207	0.908	0.906	0.918	0.903	0.424
TBV Φ_2	0.182	0.228		0.053	0.063	0.092	0.246	0.130	0.727
TBV Φ_3	0.096	0.156	-0.055		0.100	0.124	0.242	0.022	0.443
GEBV t600_I	0.960	0.956	0.201	0.047		0.989	0.955	0.965	0.278
GEBV t600_II	0.965	0.965	0.201	0.107	0.995		0.965	0.976	0.332
GEBV A	0.941	0.956	0.376	0.163	0.963	0.971		0.936	0.502
GEBV B	0.960	0.959	0.311	-0.015	0.984	0.983	0.974		0.330
GEBV C	0.468	0.527	0.749	0.382	0.411	0.454	0.636	0.502	

[1]TBV: true breeding values for yield at time 600 (t600) and logistic growth curve parameters Φ_1, Φ_2, Φ_3.

[2]GEBV: genomic breeding values for t600 estimated with method I and II (t600_I, t600_II, respectively) and for Gompertz curve parameters A, B and C.

Appendix 4.5

Correspondence between simulated QTL positions and variances with those of QTL effects mapped from variance component and association analyses

Chr.	Param.[1]	Actual QTL Position (cM)	Actual QTL Var. (%Vp)	Variance Component Analysis[2]	Association Analysis[23]
1	Φ_1	42	29.3	** 35	**19
1	Φ_2	54	32.3		
1	Φ_3	88	23.4		
2	Φ_1	5	4.6	** 6	**5
2	Φ_2	33	33.0	** 7	
2	Φ_3	49	48.9		
2	Φ_1	89	3.7		**3
3	Φ_2	7	3.5		
3	Φ_3	26	4.7	5	**3
3	Φ_2	56	3.8		
3	Φ_1	90	4.1	** 5	
4	Φ_3	10	5.9	** 6 IMP	
4	Φ_2	37	3.2	** 6	**4
4	Φ_1	70	3.3	** 7	**5
4	Φ_3	86	6.6		
5	Φ_3	31	4.6	3 DOM	
5	Φ_2	60	3.7		
5	Φ_1	77	2.5	** 5	

[1]Parameters from logistic function used to simulate data: asymptote (Φ_1), time at inflection point (Φ_2) and relative growth rate (Φ_3).

[2]The estimated % of Vp explained by each QTL is given in each analysis.

[3]All identified QTL (SNPs) were within ±5cM and a significant association with yield was identified (nominal P<0.001).

Appendix 4.6

Number of loci out of 18 simulated QTL identified by each QTL mapping method for yield and Gompertz growth model descriptors

Phenotype	Method[1,2]				
	Sire Half-sib Analysis	Dam Half-sib Analysis	Combined Sire and Dam Half-sib Analysis	Variance Component Analysis	Association Analysis
Actual Yield	9	7	12	10	6
Predicted Yield & Growth Rate	9	6	11	N/A	N/A
Gompertz parameter A	7	4	9	N/A	N/A
Gompertz parameter B	6	5	9	N/A	N/A
Gompertz parameter C	5	1	6	N/A	N/A
Combined growth model descriptors	13	8	16	N/A	N/A

[1]Dam half-sib, association, and variance component analyses of actual yield resulted in 1, 2, and 5 false positive QTL, respectively.

[2]QTL of nominal significance were also taken into account for all methods.

CHAPTER 5

Application of random regression techniques to dissect age-dependent quantitative trait loci (QTL) for growth in Blackface lambs

5.1 Introduction

As was extensively described in chapter 3, growth is a longitudinal trait that is a composite of growth rate phenotypes over time. Patterns of genetic correlations among live weights at different ages often demonstrate the trait complexity. For example, using sheep data Riggio *et al*. (2008) showed that inter-age genetic correlations for live weight, whilst strongly positive, are often different from unity, with the correlation decreasing as the time between weight measurements increases. Thus, it is likely that distinct loci act on live weights at different growth stages.

For the detection of quantitative trait loci (QTL) that are associated with growth, it would be beneficial to simultaneously analyze multiple measurements and take account of the correlation structure among measurements ordered across time. For this purpose, in chapter 3 a Gompertz growth curve was fitted to live weight information over time to map growth QTL for the curve parameters and predicted growth descriptors (Hadjipavlou & Bishop 2009). In that study, QTL were identified whose significance and variance changed over time, with QTL on distinct

chromosomes being significant at different age ranges. However, growth curves combine data only at the phenotypic level, potentially missing information arising from the different patterns of change in the genetic and environmental components of the traits.

The main objective of the current study was to utilise random regression (RR) models for chromosome-wide detection and quantification of QTL that influence longitudinal live weights, using the same live weight data from Scottish Blackface sheep that were previously analysed with the Gompertz growth model approach (chapter 3). The hypothesis underlying these analyses is that RR techniques will allow a more accurate and informative dissection of these data than the growth curve approach.

RR methodology, in principle, facilitates longitudinal trait analysis, by accounting for the covariance among trait measurements at different points in time, while allowing more flexible model fitting than a growth curve. The flexibility of RR rests on the fact that it allows separate modelling of the different time-dependent effects comprising longitudinal phenotypes of an animal. Specifically, with RR: a) a parametric curve is fitted as a fixed regression to represent the average curve of the population for live weights over time; b) random coefficient parametric curves are used to directly model each of the relevant time-dependent random effects in a mixed model. With this formulation, the parametric curve for each random effect that changes over time is thought to describe the individual's systematic time-dependent deviation, due to that effect, from the fixed curve of the population. Thus, from these random effects, time-dependent variance components can be estimated.

RR has been previously used by many studies to dissect time-dependent polygenic and maternal genetic components, along with environmental effects, for several longitudinal traits, including (but not limited to) live weights in beef cattle and sheep (e.g. Meyer 1998; Lewis & Brotherstone 2002; Fischer *et al.* 2004b; Molina *et al.* 2007), milk yield in dairy cattle (Brotherstone *et al.* 2000; Bignardi *et al.* 2009a,b), egg production in turkeys (Kranis *et al.* 2007), and repeated faecal egg count measurements in lambs (Vagenas *et al.* 2007). Yet, only a few studies have proposed RR models for identifying a QTL effect. Lund *et al.* (2002) and Macgregor *et al.* (2005) described RR approaches to analyse simulated data based on simplified biological models for animals and human populations, respectively. Both modelled a QTL as a random effect and covariance functions were used to describe the relationships between QTL effects at different ages. These two studies assessed QTL detection power of RR models fitted to longitudinal data with various time-dependence scenarios for the QTL. In addition to the QTL, Lund *et al.* (2002) simulated the same time-dependent polygenic effect for all scenarios, and this was then fitted using RR polynomials in the models tested. On the contrary, no polygenic effects were included and constant environmental effects were assumed over time in Macgregor *et al.* (2005).

Recently, Lund *et al.* (2008) extended the RR model, used previously in the simulation study reported by Lund *et al.* (2002), to allow for a genome scan to identify different time-dependent QTL, testing their methodology against univariate QTL analysis for milk yield in dairy cattle. In Lund *et al.* (2008), a time-dependent polygenic effect was included in the RR model. With the current work, the assessment of RR as an alternative technique for age-dependent QTL mapping and

description was sought by using the Blackface longitudinal live weight dataset. A further objective was to compare the RR results with longitudinal QTL mapped using the growth model analysis.

5.2 Materials and Methods

5.2.1 Data

The data used were a subset of those described in chapter 3. They comprised live weights at birth and at 4-week intervals up to 28 weeks for 788 lambs from 9 half-sib families of Scottish Blackface sheep. At least five phenotypes were available for all animals included in the analysis. The number of progeny per family ranged from 34 to 154 individuals and the animals were bred over 3 years (2001–2003). Standard records (such as parentage, day of birth, sex, birth rank) were collected. Pedigree information included 1119 animals and two generations of ancestors in addition to the lambs and their parents. Lambs were genotyped for informative microsatellite markers, as detailed in Davies *et al.* (2006) and Karamichou *et al.* (2006), on OAR 1-3, 5, 14, 18, 20 and 21. Marker map construction was explained in chapter 3 and marker order on each chromosome was summarised in Appendix 3.1. A total of 135 markers were used (range per chromosome=8-34; mean=17; median=15).

5.2.2 Random regression model

A RR model was used to fit the live weight data, first without accounting for a QTL and subsequently with the inclusion of a QTL effect. The generalized RR equation used for the full model is:

$$\mathbf{y= X\beta + Wq + Z_1u + Z_2p + Z_3c + Z_4e}$$

where **y** is a vector of observations taken at several time points for each lamb; **β** is a matrix of age-dependent fixed effects, including a fixed regression describing the population mean trajectory over time; **X** is the design matrix connecting fixed effects with records; **q**, **u**, **p** and **c** are vectors containing systematic time-dependent deviations from the fixed curve, modelled as random effects, due to allelic effects of the QTL, polygenic, permanent animal, and common environment (or litter) effects, respectively, mapped onto the data by design matrices **W, Z_1, Z_2,** and **Z_3**; **e** is the vector of residual effects, mapped onto the data by matrix **Z_4**. Vector **q** is of dimension $2N_gp_1$, where N_g is the number of animals included in the gametic matrix. Vector **u** is of dimension N_ap_2, N_a being the number of animals in the relationship matrix (i.e. pedigree). The permanent animal vector **p,** the litter vector **c** and the residual vector **e** are of dimensions N_pp_3, N_pp_4 and N_pp_5, respectively, where N_p is the number of animals with records. Parameters p_1, p_2, p_3, p_4, p_5 correspond to the number of random regression coefficients used to model the associated random effect.

Matrices **W**, **Z_1**, **Z_2** and **Z_3** are design matrices of the RR curve. The elements of these matrices are Legendre polynomials of specific order for lamb i (Φ_i). $\Phi_i=\forall_i(t^*_{ij})$, where $\forall_i$ are coefficients of the chosen Legendre polynomial for lamb i used to calculate Φ_i at age j (t^*_{ij}). The age values, which ranged from 0 to 210 days,

were standardized between -1 and +1. In cases when heterogeneous residual variance was assumed, $\mathbf{Z_4}$ was fitted as a diagonal matrix of distinct variance for each age class. The random vectors **q**, **u**, **p**, **c**, **e**, are assumed to be mutually independent and to follow multivariate normal distributions:

$\mathbf{q_i}|M, c_i \sim MVN(0, \mathbf{K_{0i}} \otimes \mathbf{Q_i}|M,c_i)$; $\mathbf{u} \sim MVN(0, \mathbf{G_0} \otimes \mathbf{A})$; $\mathbf{p} \sim MVN(0, \mathbf{P_0} \otimes \mathbf{I})$; $\mathbf{c} \sim MVN(0, \mathbf{L_0} \otimes \mathbf{I})$ and $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma^2_{ek})$, where σ^2_{ek} is the residual variance of age class k.

In the above specification, $\mathbf{K_{0i}}$, $\mathbf{G_0}$, $\mathbf{P_0}$, $\mathbf{L_0}$ are covariance matrices among random regression coefficients, **A** is the additive genetic relationship matrix and $\mathbf{Q_i}|M,c_i$ is the gametic relationship matrix of the allelic effects at the i^{th} QTL, conditional on marker data (M) and the position (c_i) on the chromosome. $\mathbf{K_{0i}}$, $\mathbf{G_0}$, $\mathbf{P_0}$, $\mathbf{L_0}$ were specified to be unstructured general covariance matrices and positive definite. The number of (co)variances among the random regression coefficients were $[p_n(p_n+1)]/2$, where $p_n \times p_n$ are the corresponding dimensions of each of the four covariance matrices**.** Thus, p_n corresponds to order+1 of the random regression polynomial fitted; e.g. fitting a polynomial of 2^{nd} order results in three random regression coefficients (coefficients of order zero, one and two) whose variances and pairwise covariances need to be estimated.

In all models, relevant identifiable fixed effects (sex, month at measurement, age of dam, year by management group, type of birth) and all two-way interactions of month at measurement with the other fixed effects were fitted. Two types of data analyses were performed. Firstly, untransformed phenotypes were analysed, fitting heterogeneous residual variances for seven time intervals. The first interval combined birth and 4-week weight records, the following five corresponded to data

for each subsequent month and the 7th interval combined 24- and 28-week weights. A fixed regression of 6th order, analogous to the population mean in single time-point analyses, was found significant and was included in models. Secondly, log-transformed phenotypes were studied, assuming homogeneous residual variances. A fixed regression of 8th order was deemed significant for transformed phenotype analyses. In both approaches, all time-dependent random effects described in the full model equation were considered.

5.2.3 Model choice, statistical testing and variance ratio construction

Prior to fitting a QTL, the likelihood ratio test (LRT) was used to test the significance of polygenic, permanent animal and litter effects across nested mixed RR models with varying order of the random regression polynomials for each effect, in order to choose the optimal "no QTL" model, as it will be referred to henceforth. The LRT statistic was calculated as twice the difference between the log likelihood of the full and the reduced model, and was assumed to have a chi squared distribution with degrees of freedom equal to the number of extra parameters estimated in the full model compared to the reduced one.

The optimal "no QTL" RR model was then used to fit a QTL effect for each chromosome. The QTL effect was then modelled with random coefficient parametric curves (described above) to allow for systematic effects of the QTL on the deviation of the animal phenotype from the expected value over time, and hence allow for changes in the QTL variance with age. LRT statistics were used to assess a) the

overall significance of the QTL effect in a RR model and b) different orders for the random regression polynomial. These analyses were performed for chromosomes 5, 14, 18 and 20. As with tests of nested "no QTL" models, the test statistic at a given location in both cases was assumed to have a naïve chi-squared distribution. Assuming a simple chi-squared distribution for the test statistic for a QTL is generally thought to be conservative at a single location in the genome, since the test statistic under the null hypothesis is expected to be distributed as a complex mixture of distributions (Self & Liang 1987; Stram & Lee 1994; Allison *et al.* 1999; Visscher 2006). In addition to this, the test statistic is conservative in the RR setup as the (co)variance terms estimated with a RR model are on the boundary of the parameter space under the null hypothesis (Meyer 1998; McGregor *et al.* 2005) and therefore the asymptotic test statistic distributions are again likely mixtures of chi-squared distributions.

All RR mixed model analyses were performed using ASReml (Gilmour *et al.* 2002). ASReml calculates the inverse of the polygenic relationship matrix **A** from pedigree information, but requires to be supplied with the inverse of the gametic relationship matrix **Q**. Loki was employed (Heath 1997) to calculate **Q**, using sire and offspring marker genotypes and pedigree information. The probability of identity by descent (IBD) between two individuals was estimated at 5 cM intervals along each chromosome across the 9 half-sib families and two generations of ancestors. Matrix inversion was achieved using the R statistical package (R Development Core Team 2006).

ASReml employs an Average Information REML procedure to maximise the restricted log likelihoods of a mixed model. One thousand iterations were specified

to allow log likelihood maximisation and model convergence. For each RR model that included the QTL effect, the likelihood of the full model was maximised every 5cM for each order of RR for the QTL effect.

For every RR analysis that achieved convergence, ASReml provided solutions for the (co)variances among RR coefficients for each time-dependent RR effect (i.e. estimation of matrices $\mathbf{K_0}$, $\mathbf{G_0}$, $\mathbf{P_0}$, $\mathbf{L_0}$) and evaluation of the fitted Legendre polynomials to each effect at each age point at which the weight was evaluated (i.e. construction of matrices $\mathbf{W}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ and $\mathbf{Z_3}$). Following Kirkpatrick *et al.* (1990), the general expression used to estimate the (co)variance matrix $\hat{\mathbf{G}}$ for each time-dependent effect separately, is the following:

$$\hat{\mathbf{G}} = \mathbf{\Phi}\, \hat{\mathbf{C}}_0 \mathbf{\Phi}^T$$

$\hat{\mathbf{G}}$ is a symmetric matrix with dimensions equal to $l \times l$, where l is the total number of age points included in the data. $\hat{\mathbf{C}}_0$ represents $\mathbf{K_0}$, $\mathbf{G_0}$, $\mathbf{P_0}$, or $\mathbf{L_0}$, for QTL, polygenic, permanent animal, and litter effects, respectively. $\mathbf{K_0}$, $\mathbf{G_0}$, $\mathbf{P_0}$, and $\mathbf{L_0}$ are symmetric matrices with dimensions equal to order+1 of the Legendre polynomial fitted for each effect ($p_1 \times p_1$, $p_2 \times p_2$, $p_3 \times p_3$, $p_4 \times p_4$, respectively). $\mathbf{\Phi}$ corresponds to the Legendre polynomials fitted for each effect, evaluated at each age point, i.e. $\mathbf{\Phi}$ is the respective matrix $\mathbf{W}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$ or $\mathbf{Z_3}$ for each effect, with dimensions $l \times p_i$.

The phenotypic (co)variance matrix for live weight is obtained by summation of the (co)variance matrices estimated for all time-dependent effects and the residual effect. In addition, correlations between ages j and k can be calculated for each effect as:

$$r_{jk} = \text{cov}(j,k)/[(\text{var}(j)\text{var}(k))^{1/2}]$$

where var(j), var(k) are the variances of ages j and k, respectively, and cov(j,k) is the covariance estimate between ages j and k. As an example, the polygenic covariance

matrix is estimated as: $\hat{\mathbf{G}} = \mathbf{Z}_1 \mathbf{G}_0 \mathbf{Z}_1^T$. The diagonal elements of $\hat{\mathbf{G}}$ correspond to the polygenic variances of live weight at each time point. The off-diagonal elements are estimates of the polygenic covariances of live weight between pairs of ages.

5.3 Results

5.3.1 "No-QTL" model choice for RR analysis of untransformed phenotypes

When untransformed live weights were analysed using RR, systematic model search prior to fitting a QTL led to a great number of models fitted. Likelihood estimates for the majority of "no QTL" models fitted are summarised in Appendix 5.1. Table 5.1 shows the LogL estimates from one group of "no QTL" models that was chosen for further study. In this model group, a polygenic effect was fitted with a Legendre polynomial of order 1 and the polynomial order varied for the two environmental effects. In Table 5.1, nested likelihood ratio testing can only be performed across rows or down columns, and the degrees of freedom associated with each order of fit are 1, 3, 6, 10, 15 and 21, respectively.

Table 5.1 Log-likelihood estimates for "no QTL" RR models for untransformed live weights over time

Random regression polynomial order[1][2]		Permanent animal effect					
		0	1	2	3	4	5
Litter effect	0	-6300	-6336	-6132	-6131	NC[3]	NC
	1	-6303	-6394	-6193	-6194	NC	NC
	2	-6118	-6225	-6190	-6163	-6160	-6180
	3	-6130	-6243	-6179	-6194	-6202	NC
	4	-6173	-6286	-6177	-6199	-6232	NC
	5	-6219	-6332	-6196	-6211	-6250	NC

[1]A polygenic effect with a random regression polynomial of order 1 was fitted in each model.

[2]RR polynomial order is equivalent to order+1 random coefficients estimated.

[3]NC=Model fit did not converge.

As can be seen in Table 5.1, some higher order models clearly did not converge. In other cases, the LogL did not improve when the order of either litter or permanent animal effect (or both) increased. This indicated that, for these cases, the likelihood had probably failed to reach a global maximum and converged at local maxima instead. Consideration of the above resulted in the rejection of all models with either of the two environmental effects fitted to order higher than 3. Inspection of the predicted variance estimates for each component in the remaining models led to the choice of a model with a 2nd order RR polynomial fitted for both environmental effects as the optimal "no QTL" model. Profiles of the proportion of phenotypic variance for each effect, predicted using this model, are shown in Figure 5.1. Plots are discontinuous because heterogeneous residual variances were fitted for the seven time intervals. In this figure, the predicted polygenic heritability seems somewhat overestimated, in particular at the edges of the time range.

Other models with better likelihood estimation than the chosen one were avoided since the variance partitioning among the random effects fitted in these

models did not seem plausible. For example, a model with zero and second order polynomial fitted for litter and permanent animal effects, respectively, led to substantially higher variance estimation for litter than for polygenic effect (Appendix 5.2).

Genetic correlations between actual live weights at different ages were estimated using the (co)variance predictions from this model, and are given in the lower diagonal of Table 5.2. These were in agreement with previous estimates from bivariate analyses of a more extensive dataset that included the phenotypes analysed in the current study (Riggio *et al.* 2008). Genetic correlations decreased as the time between the measurements increased, as was also observed previously (Fischer *et al.* 2004b; Fischer *et al.* 2006; Riggio *et al.* 2008).

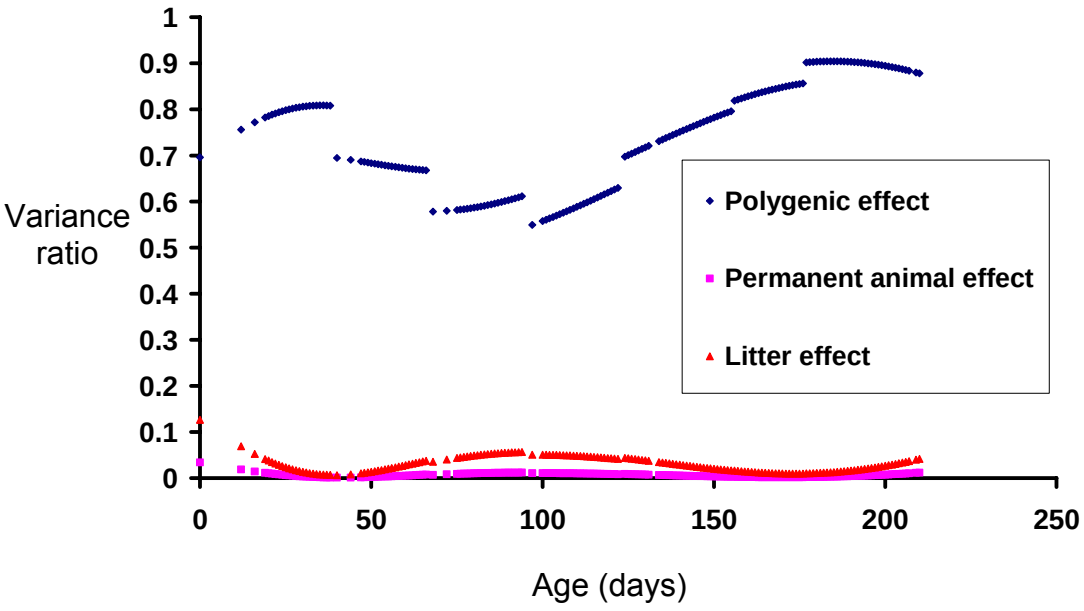


Figure 5.1 Profiles of the proportions of phenotypic variance for polygenic, litter and permanent animal effects predicted using chosen RR model without fitting QTL effects.

Table 5.2 Genetic correlations between live weights at various ages predicted using chosen RR models without fitting QTL effects[12]

Day	0	28	56	84	112	140	175
0	1	0.60	0.42	0.34	0.30	0.29	0.29
28	0.83	1	0.98	0.94	0.91	0.84	0.59
56	0.51	0.91	1	0.99	0.97	0.91	0.65
84	0.24	0.75	0.96	1	0.99	0.95	0.70
112	0.05	0.60	0.88	0.98	1	0.98	0.79
140	-0.07	0.50	0.82	0.95	0.99	1	0.89
175	-0.17	0.41	0.76	0.92	0.98	0.99	1

[1]Correlations determined from modelling untransformed and transformed phenotypes given in lower and upper diagonal, respectively.

[2]The respective optimal "no QTL" model was used in each analysis: RR order 1 for polygenic effect and order 2 for permanent animal and litter effects for untransformed data; RR order 2 for polygenic and litter effect for transformed data.

5.3.2 Full model choice and results for RR analyses of untransformed phenotypes

Full RR models were fitted separately for OAR 5, 14, 18 and 20. OAR14, OAR18, and OAR20 were explored because age-dependent QTL were previously found on these, using live weights and growth rates predicted from the Gompertz growth curve (chapter 3; Hadjipavlou & Bishop 2009). OAR5 was also studied as a potential "negative control" for the RR model fitting, since no OAR5 growth QTL were identified using the growth model approach (chapter 3).

5.3.2.1 OAR14

The log likelihoods of different models and test statistics for an OAR14 QTL are shown in Table 5.3. A first order Legendre polynomial for the QTL effect in the full model resulted in a test statistic estimate against the "no QTL" model that was not significant. This coincided with the QTL variance not being estimable (or estimated to be close to zero) at certain chromosomal positions. A significant QTL was supported by models with polynomial order of 2 ($P<0.05$) or higher ($P<0.01$), although the most likely position of the QTL was different depending on the QTL order fitted (50, 105, 115 cM with orders 2, 3, 4, respectively). Further, a second test statistic for increasing the order of the RR polynomial is given in the table, and this provided justification for fitting a QTL with a 3rd order but not 4th. The same trend for a highly significant QTL fitted with 2nd RR order or higher was obtained if the permanent animal effect was omitted from the above model (Appendix 5.3).

Table 5.3 Log-likelihood estimates and test statistics for OAR14 QTL effect fitted in chosen "no QTL" RR model using RR polynomials of varying order

QTL RR order in model	LogL	Test statistic when compared to no QTL model[1]	Test statistic of increasing QTL order[1]
1	-6189	2 (3 d. f.)	NA
2	-6184	13(6 d. f.)*	11 (3 d. f.)**
3	-6159	62 (10 d. f.)**	49 (4 d. f.)**
4	-6156	67 (15 d. f.)**	5 (5 d. f.)

[1]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

Figure 5.2 shows the proportion of phenotypic variance partitioned to the QTL effect across time in a model with a QTL RR of order 3. Since the estimated variance ratios across time differ only slightly, it is difficult to come to a conclusion regarding the trend in variance change across time. A time-dependence trend may be speculated, with a maximum effect close to 65 days, but it is masked by inflated variance estimates at the end age points, which possibly arise as a consequence of prediction using regression. Additionally, the variance ratio profiles over time for the other time-dependent effects (Figure 5.3) indicated overestimation of the polygenic heritability over time (Figure 5.3a) and confounding of the permanent animal and QTL effect (Figure 5.3b), presumably because both terms are correlated with an animal's Mendelian sampling term.

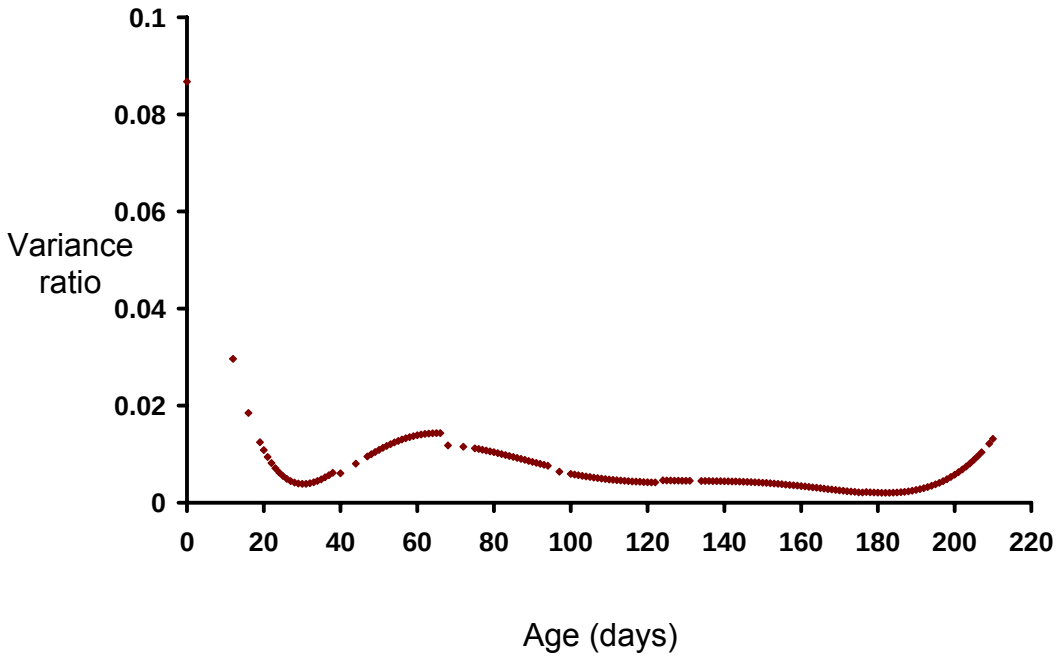
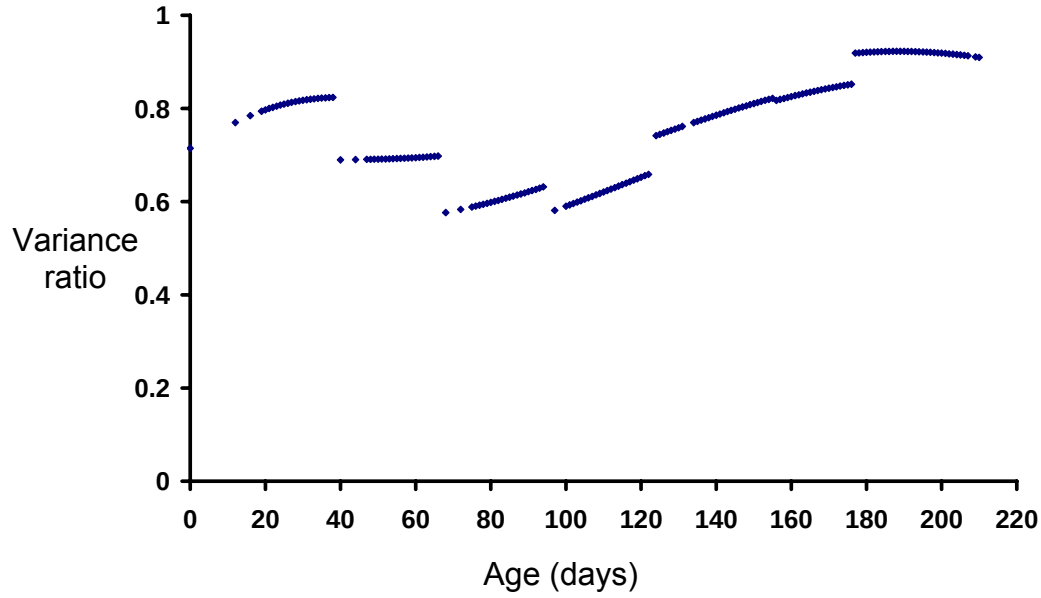
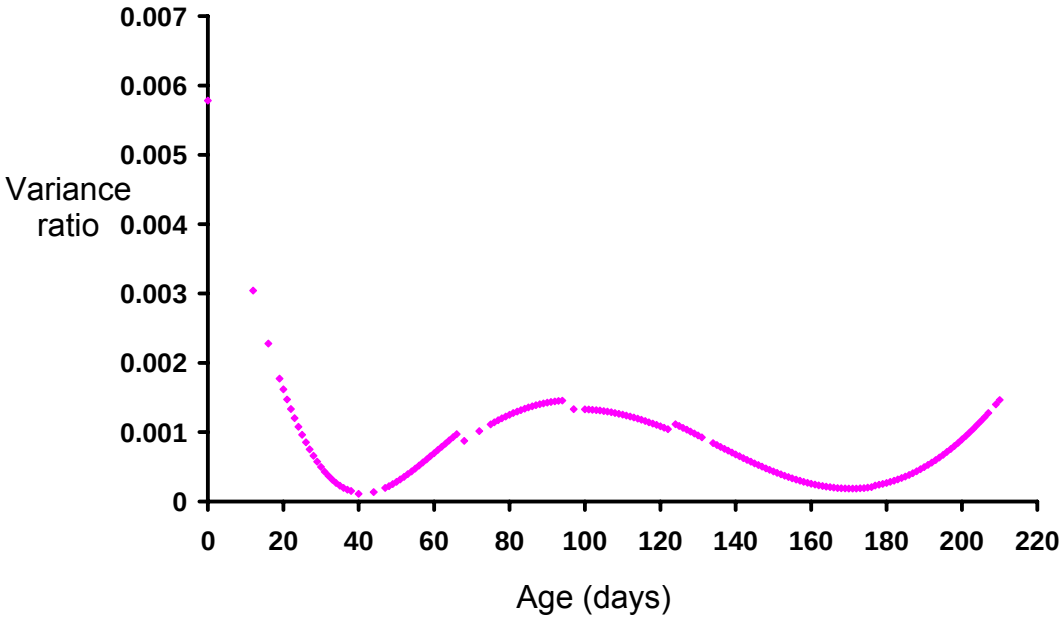


Figure 5.2 Proportion of phenotypic variance explained by the QTL across ages on OAR14, predicted by the chosen full RR model.

a)


1
0.8
0.6
0.4
0.2
0
Variance ratio
0
20
40
60
80
100
120
140
160
180
200
220
Age (days)

b)


0.007
0.006
0.005
0.004
0.003
0.002
0.001
0
Variance ratio
0
20
40
60
80
100
120
140
160
180
200
220
Age (days)

c)

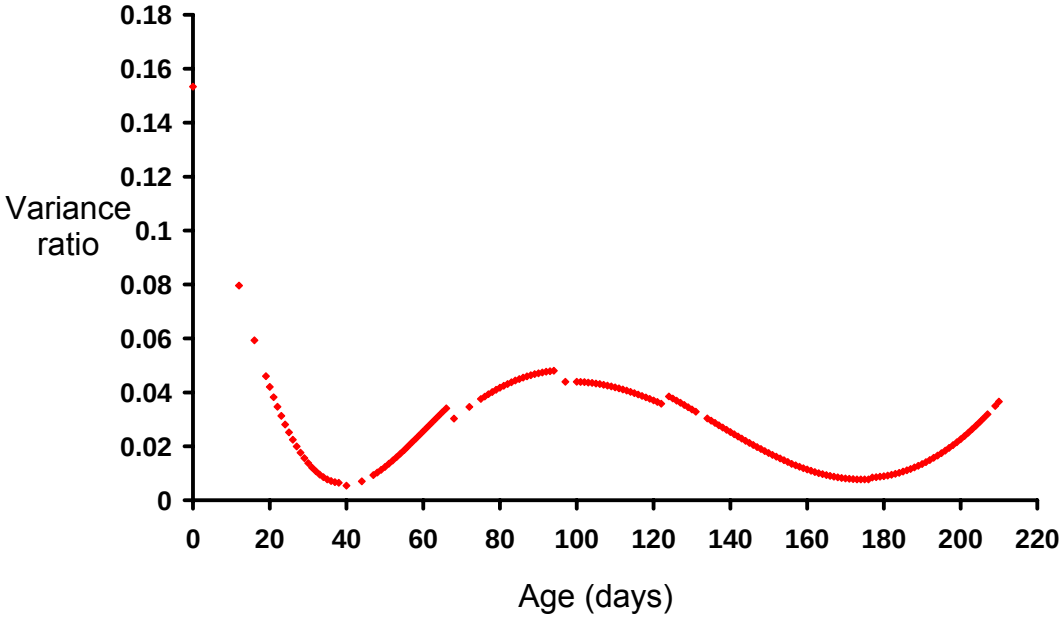


Figure 5.3 a) Polygenic heritability and proportion of phenotypic variance explained by b) the permanent animal and c) litter effects across ages on OAR14, predicted by the chosen full RR model.

5.3.2.2 OAR5, OAR18 and OAR20

A similar pattern of test statistic results to those obtained for OAR14 was seen when the chosen "no QTL" RR model was used to fit a QTL effect using IBD information for each of the other 3 chromosomes studied (Appendix 5.4). A 2nd order RR polynomial for the QTL effect resulted in convergence and provided evidence for significant QTL ($P<0.05$) at different positions on each chromosome, although no significant differences were found among LogL at distinct positions on each of the chromosomes. The significance for including a QTL effect in the model increased when the RR order increased to 3rd and 4th order. A 3rd order RR for the QTL was supported by LRT. The QTL variance profiles across time on the 3 chromosomes

resembled that obtained for the QTL on OAR14 (Appendix 5.5). The QTL variance ratio again appeared to increase at intermediate ages with edge effects present as before. Confounding was suggested among the fitted effects, and in particular between the QTL and permanent animal effects.

5.3.3 "No QTL" and full model choice and results for RR analyses of transformed phenotypes

Following analyses of untransformed data, RR analysis of natural logarithm-transformed weights was explored for the same chromosomes. Again, for "no-QTL" model choice, different orders of the polygenic, permanent animal effect and litter effects were systematically examined. The permanent animal effect was not significant based on LRT of nested models or led to no convergence when fitted with 1st RR order or higher (Appendix 5.6). As can be seen in Table 5.3, for transformed trait RR analyses, the LogL estimates obtained for nested models improved as the order of the effects increased, indicating better convergence properties than the models fitted for untransformed data analysis. Models with 3rd RR order or higher for either effect resulted in no convergence, similar to what was observed in the model fitting process of untransformed phenotypes. A model with 2nd RR order for both the polygenic and litter effects was the model best describing the data using LRT (Table 5.3). The variance ratios over time for polygenic and litter effects predicted from the chosen "no QTL" model are plotted in Figure 5.4, and seem closer to expectation than those obtained from RR analyses of untransformed data. Polygenic heritability appeared to increase at early ages and to remain relatively

constant from 80 days onward. The litter effect was predicted to explain 20% of the phenotypic variance for transformed live weight at intermediate ages and to decrease from 130 days onward.

Genetic correlations between live weights at different ages were estimated using the (co)variance predictions from the optimal "no QTL" model, and are given in the upper diagonal of Table 5.2. On average, slightly higher correlations were found between the most distant age points for transformed compared to untransformed weights. Overall, though, a similar trend of correlations decreasing as the interval between time points increases was observed.

To dissect a QTL on OAR14, a zero order RR for the QTL was first fitted in the optimal "no-QTL" model. Even though the model converged across the chromosome except at 60 and 65cM, the QTL variance was not estimable at most positions with convergence. In the region where the variance was estimated (85-120cM), the QTL was not significant. Increasing the QTL order to 1, led to convergence problems at all positions. When a QTL was included in reduced order "no QTL" models, only a zero order for the QTL allowed convergence. In suboptimal models, the QTL variance was still not estimable at many positions, and where it was estimable, the QTL was not significant (Appendix 5.7). Analogous outcomes were obtained when OAR5 and OAR18 were studied, whereas for OAR20, when a zero order RR was fitted for the QTL in the optimal "no QTL" model, the QTL variance was not estimable at any chromosomal position.

Table 5.3 Log-likelihood estimates for "no QTL" RR models for transformed live weights over time

Random regression polynomial order[1]		Polygenic effect				
		0	1	2	3	4
Litter effect	0	7426	7617	7851	NC[2]	NC
	1	7564	7622	7857	NC	NC
	2	7710	NC	7871	NC	NC
	3	NC	NC	NC	NC	NC
	4	NC	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

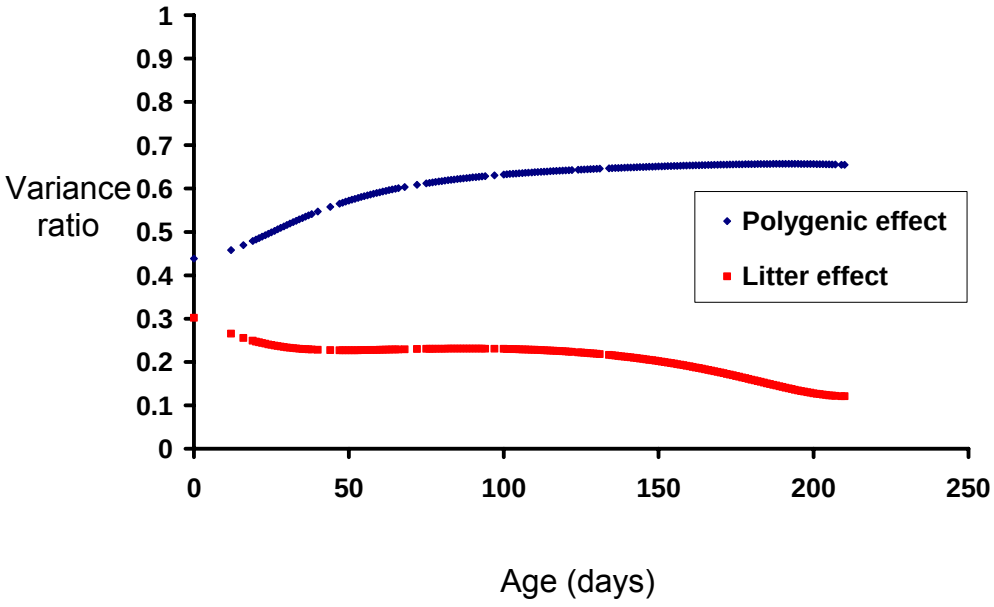


Figure 5.4 Profiles of the proportion of phenotypic variance explained by the polygenic and litter effects, predicted using chosen RR model without fitting QTL effects.

5.4 Discussion

This study sought to test RR methodology as an alternative, more flexible approach to growth curve modelling (chapter 3) with the aim of identifying age-dependent QTL for growth using live weight information across time. This work is, to my knowledge, the first to use RR models to pursue mapping of QTL whose effects on growth change with time. A recent study used RR models to find age-dependent QTL for milk yield in dairy cattle (Lund *et al.* 2008). The current procedure differed from that of Lund *et al.* (2008) in that a more complete model was explored with the aim of including and describing variance components that are relevant to longitudinal phenotypes. Specifically, in this study: a) in addition to QTL and polygenic effects, time-dependent permanent animal and common environment (litter) effects were investigated in the RR setup; b) heterogeneous residual variances were included in analyses of untransformed live weights, whilst transformed phenotypes were also studied separately; and c) a systematic two-stage strategy was followed to choose the full RR model that best described the trait.

The first step of the model choice approach resulted in the evaluation of numerous RR models with different combinations of relevant time-dependent random effects prior to fitting a QTL. LRT statistics, along with prediction properties of the different models, were used to assess the "no QTL" models. In the second step of the model choice process, two test statistics based on LRT were used to assess a) overall QTL significance and b) significance of varying RR orders for the QTL effect in the chosen "no QTL" model. All test statistics were assumed to follow a naïve chi-squared distribution with degrees of freedom equal to the difference in number of

(co)variances estimated in the two models compared in each case. However, assuming the above distribution and estimations of the test statistic is conservative; testing is done at the boundary of the parameter space when evaluating whether additional RR coefficients have zero variance (Meyer 1998). In analogous conditions, it has been shown that the LRT criterion follows a mixture of chi-squared distributions (Self & Liang 1987). Even so, a conservative test statistic was used in this study since the mixture of chi-squared distributions is not known for longitudinal QTL models (Lund et al. 2008), although some proposals have been made for simple scenarios (Macgregor *et al.* 2005).

Even with the systematic strategy explained above, model choice proved complex, as had been previously observed in other studies using RR modelling (e.g. Meyer 1998; Meyer 1999; Fischer & Van der Werf 2002). Formal statistical testing was not sufficient to identify the optimal RR order for the random effects included in the "no QTL" model. This was particularly true for the analysis of untransformed phenotypes, in which some models appeared to have converged at local maxima. This possibility only became apparent because sequential fitting of nested models was pursued, thus, making comparisons of LogL possible. As an approach that would assist model choice, and in addition to taking into account the LogL estimation, the prediction properties were also considered for each of the "no QTL" models examined. This appeared useful, and has previously been advocated by Meyer (1998) and Fischer & Van der Werf (2002). Yet, this still remains a subjective criterion based on prior expectation and would require further investigation in the future.

In an attempt to alleviate potential model overparameterisation, which may have led to problems in achieving LogL convergence at a global maximum in the

aforementioned models, RR analyses of transformed live weights were performed. This resulted in the reduction of number of parameters estimated in the model, since homogeneous residual variance could be assumed. Better convergence properties were indeed observed for the "no QTL" model evaluation when transformed measurements were analysed.

Estimation of genetic correlations for live weight over time, using the chosen "no QTL" models for untransformed and transformed phenotype analyses, resulted in moderate to strong correlations being observed. The trends observed agreed with previous findings from bivariate analyses of the same dataset (Riggio *et al.* 2008), and confirmed that correlations decrease as the time interval between live weight measurements increases (Fischer *et al.* 2004b; Riggio *et al.* 2008).

Even though the estimated genetic correlations allowed some confidence on the chosen "no QTL" models, taken as a whole, inconclusive results were obtained with the two separate RR analyses of Blackface live weights upon inclusion of QTL as a time-dependent random effect in the models. RR modelling of untransformed live weights led to apparent detection of significant age-dependent QTL on OAR5, OAR14, OAR18 and OAR20. This would have seemed promising had it not been for a) almost identical patterns of test statistics for a QTL and b) similar QTL variance trajectories over time for the QTL on all chromosomes studied. These outcomes seemed implausible based on the observed genetic correlation pattern over time and did not agree with the QTL identified on these chromosomes using the Gompertz growth curve approach (chapter 3). Additionally, substantial overestimation of polygenic heritability was found in all analyses of untransformed live weights. Finally, the analysis of transformed phenotypes did not provide additional

information, as convergence problems arose when the QTL effect was included in the chosen "no QTL" model with RR order of 1 or higher.

Despite the inconclusive results, this study presented some lessons with respect to model choice. In RR analyses for QTL detection, it seems necessary to carefully choose a suitable "no QTL" model that adequately describes each time-dependent random effect prior to fitting the QTL effect in the model. If an overparameterised "no QTL" model is chosen, as may have been the case in the analysis of live weights in this study, estimation of the random regression parameters for the QTL may result in implausible outcomes (as for untransformed phenotypes) or may be impaired (as for transformed live weights). Moreover, if in the chosen model the variance that is attributable to the QTL is captured by another random term, the QTL effect description would be severely diminished. In some models explored in this study, confounding was particularly apparent between the QTL and permanent animal effect. This may be expected because both terms will be correlated with the Mendelian sampling terms in the current data structure. In general, models that included at least the same order for QTL as for permanent animal effect seemed to allow inclusion of the QTL effect, even with incomplete variance partitioning between the two effects. Finally, variance predictions over time from some full models fitted indicated partial confounding between the QTL and polygenic effects.

In addition to RR model choice influencing convergence and variance partitioning, it is important to bear in mind that a limited phenotype dataset was available for RR analyses in this study both because of the experimental design and to ensure sufficient observations per lamb. Thus, it is difficult to resolve whether the uncertainty in dissecting longitudinal QTL effects with this approach arises from data

and population structure constraints, demonstrated for other time dependent variance components in previous studies (Meyer 2002; Fischer & Van der Werf 2002), or inherent properties of RR methodology, as reported earlier (e.g. Meyer 1998; Meyer 1999; Fischer *et al.* 2004a,b). Specifically, in the current study, inflated variance predictions were observed for all random effects at the end-points of the time range studied. This inflation seems most likely to be a result of prediction using regression, with endpoint predictions having greater sampling variance. Analogous outcomes were seen previously (Jamrozik & Schaeffer 1997; Van der Werf *et al.* 1998; Meyer 1999; Fischer *et al.* 2004a). In particular to QTL detection, caution is advocated due to this method limitation, as it does not allow unambiguous assessment of increased QTL variance observed at early or late ages. In addition, this makes it difficult to evaluate variance changes along the entire time range. Thus, even if the apparently significant QTL detected on OAR5, 14, 18, and 20 were to be accepted, no clear conclusions can be made with the current RR structure regarding the actual time range of maximum QTL expression.

Overall, with the given data structure and findings, the RR approach seemed to have poorer ability to describe growth QTL over time compared to the growth model approach previously used to dissect the expression of age-dependent QTL for growth rates and live weights, using the same dataset (chapter 3) from Scottish Blackface sheep.

In order to further address unresolved issues from the current RR study, a simulation study was subsequently conducted and is described in chapter 6. This study aims at examining the model choice procedure followed in the above study, variance partitioning and confounding between time-dependent components in the

RR framework, along with the ability of the method to accurately predict the pattern of changes in the variance trajectory over time for age-dependent growth QTL.

5.5 Conclusions

Random regression techniques have been previously proposed for dissecting time-dependent polygenic and environmental effects, and could be potentially applied to identify and describe the expression of age-dependent QTL for a longitudinal trait. This study explored a systematic procedure for selecting random regression models that include a QTL as a time-dependent random effect, in addition to polygenic and environmental effects. Longitudinal live weights were analysed, with microsatellite marker information being used to calculate gametic relationship matrices. Genetic correlations between phenotypes at different time points were estimated from RR models and decreased as the time between weight measurements increased, as was observed previously. When untransformed phenotypes were studied, RR models with 2^{nd} order or higher fitted for the QTL effect detected apparently significant time-dependent QTL for growth in Blackface sheep on OAR5, OAR14, OAR20, and OAR20. However, the choice of RR models with relevant variance components and the appropriate order for the random coefficient parametric curves proved generally complex, for models both with and without QTL. Moreover, model convergence often seemed to occur at local maxima. With the chosen full models, all QTL had similar variance ratios over time, and inflated polygenic heritability was observed in all analyses of untransformed weight. These findings did not agree with the pattern of differential expression for age-dependent QTL,

identified on OAR14, OAR18, and OAR20 by analysing the same dataset using a growth model approach (chapter 3). RR analyses of transformed live weights resulted in better convergence properties of models that included polygenic and environmental effects, only. However, upon inclusion of a QTL effect, convergence problems arose. Thus, implementations of the RR method on two separate forms of the data did not allow conclusive results on the statistical significance for QTL on different chromosomes. With the given data structure, the RR approach seemed to have poorer ability to describe growth QTL over time than the previously applied growth model approach. Further work is required to assess the ability of RR models to detect growth QTL with different patterns of age-dependence.

Appendix 5.1

Log-likelihood estimates for groups of "no QTL" RR models fitted to untransformed live weights over time for polygenic effect only, polygenic and permanent animal or polygenic and litter effects, and all three effects. The given RR polynomial order is equivalent to (order+1) random coefficients estimated for each effect. NC is an abbreviation for no convergence of a fitted model.

Random regression polynomial order	Polygenic effect
0	-6817
1	-6420
2	-6486
3	-6903
4	-7473

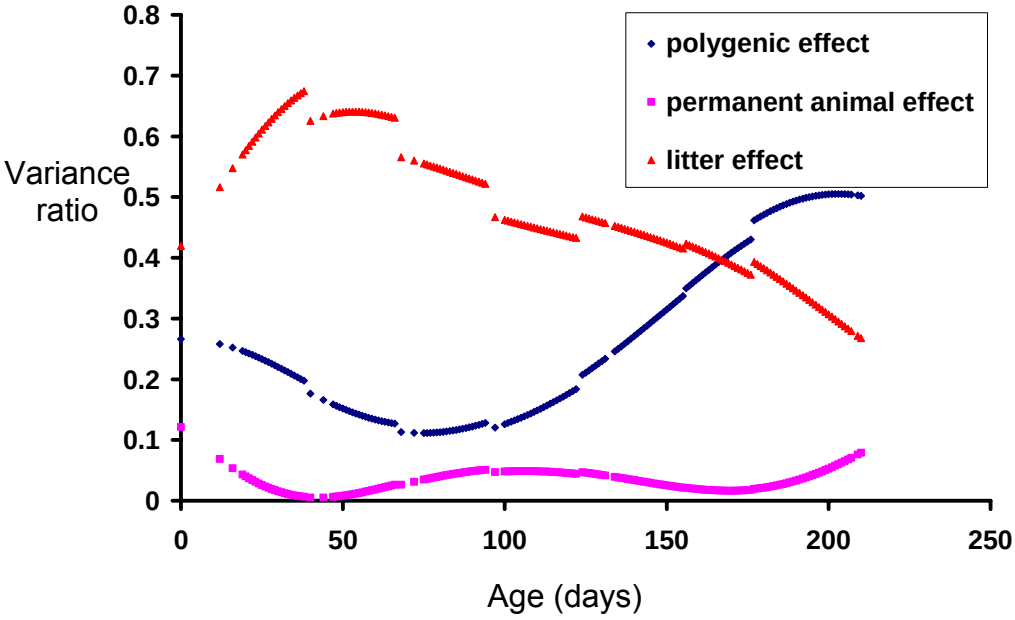
Random regression polynomial order		Permanent animal effect				
		0	1	2	3	4
Polygenic effect	0	-6819	-6325	-6173	-6324	-6556
	1	-6315	-6420	-6216	-6218	NC
	2	-6155	-6214	-6482	-6472	-6459
	3	-6309	-6214	-6500	-6905	-6910
	4	-6532	NC	-6499	-6952	NC

Random regression polynomial order		Litter effect				
		0	1	2	3	4
Polygenic effect	0	-6800	-6487	-6393	-6478	-6587
	1	-6343	-6395	-6226	-6245	-6288
	2	-6259	-6340	-6443	-6468	-6461
	3	-6498	-6575	-6697	-6846	-6898
	4	-6752	-6760	-6887	NC	NC

Random regression polynomial order (per effect)			LogL
Polygenic	Litter	Permanent animal	
0	0	0	-6799
0	1	0	-6487
1	0	0	-6300
0	0	1	-6313
1	1	0	-6303
1	0	1	-6336
0	1	1	-6304
0	2	0	-6393
2	0	0	-6134
0	0	2	-6160
2	2	0	-6132
2	0	2	-6241
0	2	2	-6143
1	2	0	-6118
2	1	0	-6103
0	1	2	-6113
0	3	0	-6478
3	0	0	-6280
0	0	3	-6306
3	3	0	-6279
3	0	3	-6479
0	3	3	-6280
1	3	0	-6130
3	1	0	-6174
0	1	3	-6168
1	1	3	-6194
1	3	1	-6243
3	1	1	-6176
2	1	3	-6327
1	2	3	-6163
3	2	1	-6168
1	1	1	-6394
1	2	1	-6225
1	2	2	-6190
2	1	1	-6178
2	2	1	-6195
2	1	2	-6324
1	1	2	-6193
2	2	2	-6440
2	3	2	-6463
3	2	2	-6438
3	2	3	-6685
2	3	3	-6468
3	3	2	-6498
2	2	3	-6431
3	3	3	-6846

Appendix 5.2

Profiles of the proportion of phenotypic variance explained by each effect fitted in a "no QTL" RR model with 1st order for polygenic, 2nd order for permanent animal, and zero order for litter effects.



Appendix 5.3

Test statistics for OAR14 QTL effect fitted in RR model with 1st and 2nd RR orders for polygenic and litter effects, respectively.[1]

QTL RR order in model	Test statistic when compared to no QTL model[2]	Test statistic of increasing QTL order[2]
1	2 (3 d.f.)	NA
2	82 (6 d.f.)**	80 (3 d.f.)**
3	130 (10 d.f.)**	48 (4 d.f.)**
4	118 (15 d.f.)**	-

[1]LogL=-6226 for "no QTL" model (1st and 2nd RR orders for polygenic and litter effects, respectively).

[2]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

Appendix 5.4

Test statistics for OAR5, OAR18, OAR20 QTL effect fitted in chosen "no QTL" RR model using RR polynomials of varying order.

OAR5

QTL RR order in model	Test statistic when compared to no QTL model[1]	Test statistic of increasing QTL order[1]
1	0.18 (3 d.f.)	NA
2	16 (6 d.f.)*	16 (3 d.f.)**
3	66 (10 d.f.)**	50 (4 d.f.)**
4	70 (15 d.f.)**	4 (5 d.f.)

[1]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

OAR18

QTL RR order in model	Test statistic when compared to no QTL model[1]	Test statistic of increasing QTL order[1]
1	2 (3 d.f.)	NA
2	14 (6 d.f.)*	12 (3 d.f.)**
3	62 (10 d.f.)**	48 (4 d.f.)**
4	69 (15 d.f.)**	7 (5 d.f.)

[1]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

OAR20

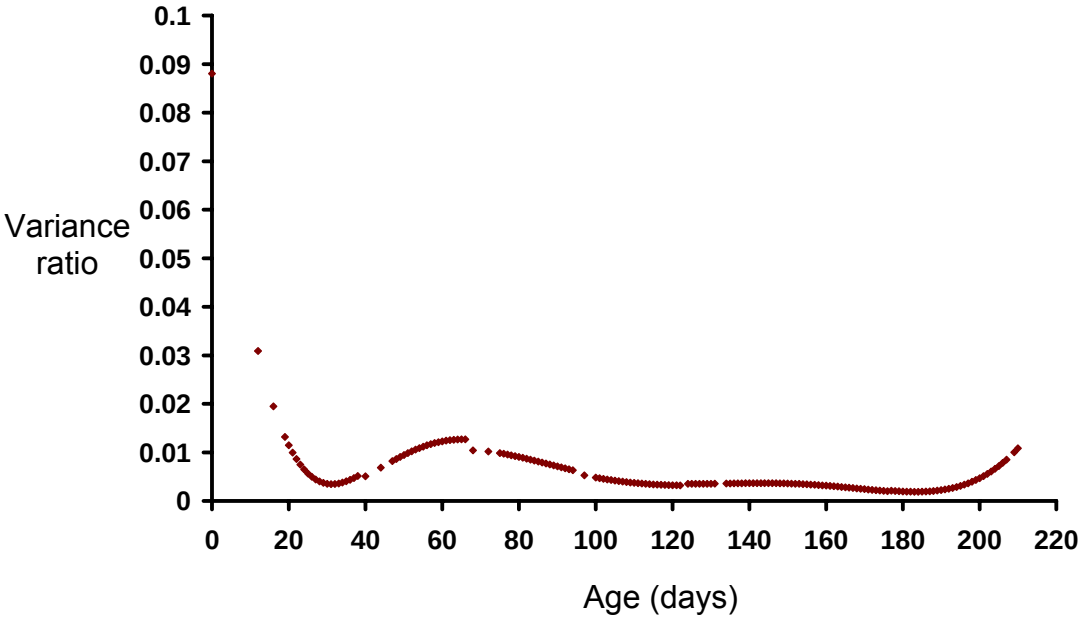
QTL RR order in model	Test statistic when compared to no QTL model[1]	Test statistic of increasing QTL order[1]
1	2 (3 d.f.)	NA
2	14 (6 d.f.)*	12 (3 d.f.)**
3	70 (10 d.f.)**	56 (4 d.f.)**
4	73 (15 d.f.)**	3 (5 d.f.)

[1]QTL effect: *significant (P<0.05); ** highly significant (P<0.01)

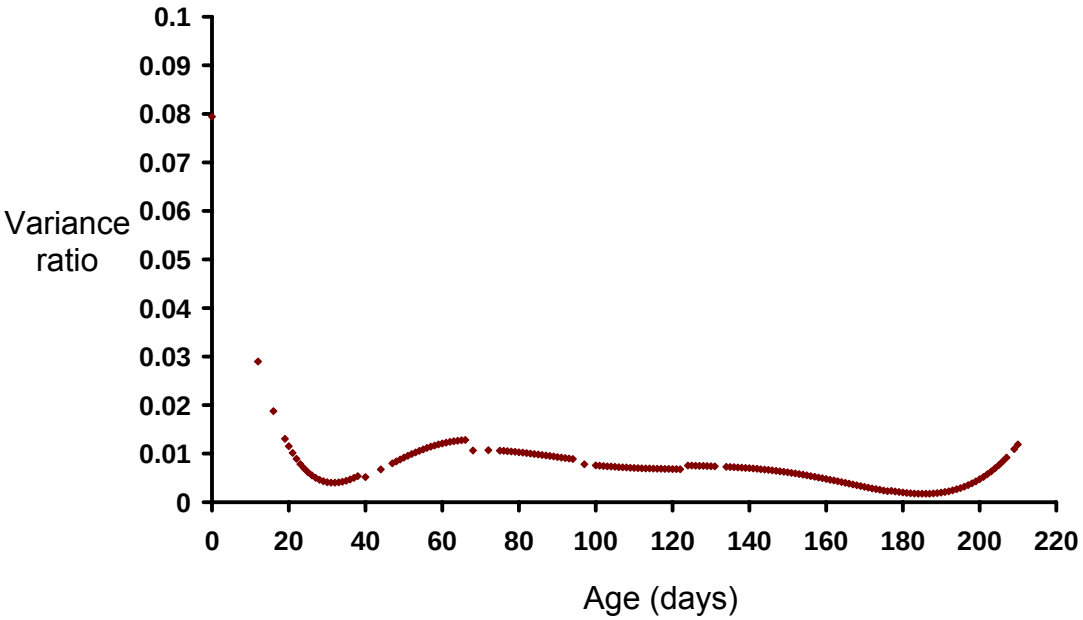
Appendix 5.5

Proportion of phenotypic variance explained by QTL across time on OAR5, OAR18, OAR20, obtained when a 3rd RR order is fitted for the QTL in the chosen "no QTL" model.

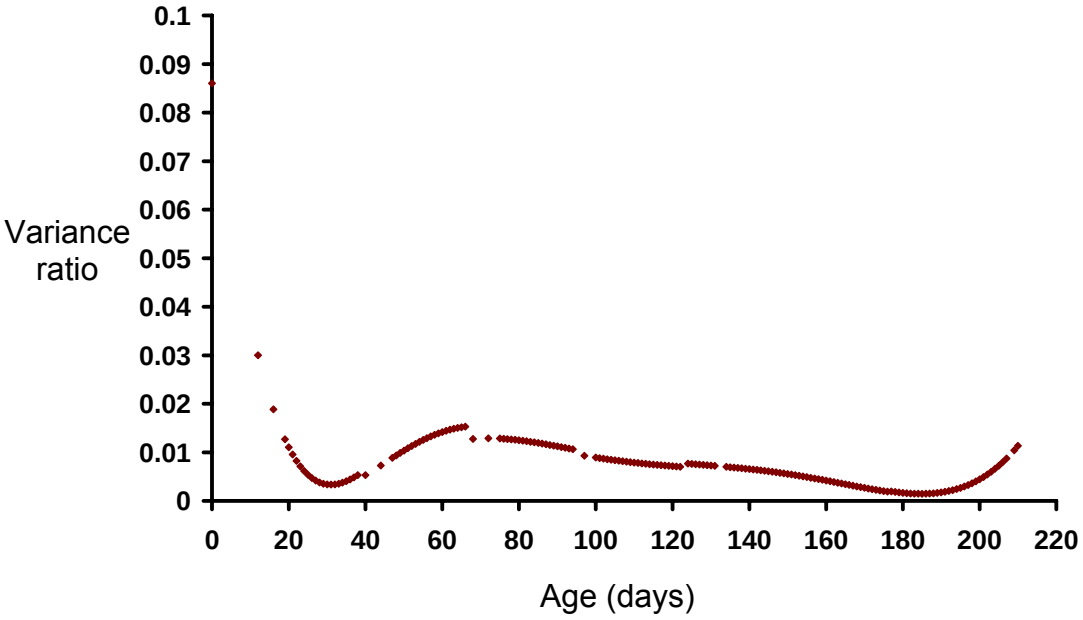
OAR5



OAR18



OAR20



Appendix 5.6

Log-likelihood estimates for additional groups of "no QTL" RR models fitted to transformed live weights over time for polygenic effect only, polygenic and permanent animal effects, and all three effects. The given RR polynomial order is equivalent to (order+1) random coefficients estimated in each model. NC is an abbreviation for no convergence of a fitted model.

Random regression polynomial order	Polygenic effect
0	7411
1	7600
2	7838
3	NC

Random regression polynomial order		Permanent animal effect	
		0	1
Polygenic effect	0	7412	NC
	1	7560	NC

Random regression polynomial order			LogL
Polygenic effect	Litter effect	Permanent animal effect	
0	0	0	7427
0	1	0	7565
1	0	0	7617
0	0	1	NC
1	1	0	7622
1	0	1	NC
1	0	2	NC
0	1	1	NC
1	1	1	NC
1	1	2	NC
2	0	0	7851
0	0	2	NC
2	2	0	7861
2	2	1	NC
2	2	2	NC
2	0	1	NC
2	0	2	NC
1	2	0	NC
2	1	0	NC
0	1	2	NC
2	1	2	NC
2	1	1	NC

Appendix 5.7

Inclusion of a QTL effect of varying orders in statistically suboptimal "no QTL" models fitting transformed live weights over time to identify OAR14 QTL.

Random regression polynomial order			Model convergence	LogL[1]	QTL variance
QTL effect	Polygenic effect	Litter effect			
0	1	1	Yes	7622 (N.S.)	Estimable at 55-65cM; 100-125cM
0	2	1	Yes (not at 60-65 cM)	7857 (N.S.)	Estimable at 85-125 cM
1	1	1	Only at 145 cM	7622 (N.S.)	Not estimable
1	2	0	No	-	-
1	2	1	No	-	-
2	0	-	No	-	-
2	1	0	No	-	-

[1]N.S.: QTL effect not significant based on LRT against respective "no QTL" model

CHAPTER 6

Exploration of random regression models to describe different time-dependent QTL trajectories affecting growth traits in a simulated sheep population

6.1 Introduction

As was also described in chapter 5, random regression (RR) methodology is, in principle, an efficient technique for analysing traits that vary with time and describing their time-dependent variance components. RR modelling accounts for the covariance among trait measurements at different points in time, without imposing a narrow phenotypic structure over time. Specifically, with RR: a) a parametric curve is fitted as a fixed regression to represent the average curve of the population for live weights over time; b) random coefficient parametric curves are used to directly model each of the relevant time-dependent variance components in a mixed model. With this formulation, the parametric curve for each random effect that changes over time is thought to describe individual systematic time-dependent deviations, due to that effect, from the fixed curve of the population.

Random regression (RR) models were previously applied on longitudinal weights in different species to estimate time-dependent polygenic (e.g. Meyer 1998; Meyer 1999; Fischer *et al.* 2004b; Lambe *et al*. 2006), along with maternal genetic and environmental effects (e.g. Lewis & Brotherstone 2002; Fischer *et al.* 2006;

Molina *et al.* 2007). RR approaches have also been used to study other longitudinal phenotypes, such as milk yield in dairy cattle (e.g. Brotherstone *et al.* 2000; Bignardi *et al.* 2009a,b), egg production in turkeys (Kranis *et al.* 2007), and repeated faecal egg count measurements in lambs (Vagenas *et al.* 2007).

However, only a few studies have proposed RR models for quantitative trait loci (QTL) mapping (Lund *et al.* 2002; Macgregor *et al.* 2005; Lund *et al.* 2008). Lund *et al.* (2002) described RR approaches for animals and Macgregor *et al.* (2005) for human populations. Both conducted simulation studies to assess QTL detection power of RR models fitted to longitudinal data with different time-dependence scenarios for the QTL. In both studies, the QTL was modelled as a random effect using covariance functions of age. However, in both simulations, a rather simplified biological model was used. In addition to the QTL, Lund *et al.* (2002) only simulated a time-dependent polygenic effect for all scenarios, and the same RR polynomial order used to simulate the effect was fitted to account for the polygenic effect in the models tested. No polygenic effects were included and constant environmental effects were assumed over time by Macgregor *et al.* (2005). Recently, Lund *et al.* (2008) extended the RR model of the earlier simulation study, to allow for a genome scan and time-dependent QTL mapping for milk yield in dairy cattle. Their main objective was to test the RR methodology against univariate QTL analysis. In addition to examining varying RR orders for the QTL, Lund *et al.* (2008) included a time-dependent polygenic effect of 3rd RR order in all models.

Overall, RR studies of actual and simulated phenotypes have indicated that this approach allows flexible model fitting for longitudinal traits, mainly because it results in a reduced number of parameters estimated than a full multivariate analysis,

and it performs significantly better than a repeatability model. In particular for time-dependent QTL mapping, Macgregor *et al.* (2005) and Lund *et al.* (2008) showed that when there was substantial QTL variance increase over time, even a first degree RR fit increased the power to detect the QTL. Also, modelling the QTL effect using a RR polynomial on age could potentially allow prediction of the QTL variance pattern over time (Lund *et al.* 2008). However, even with simple simulation structures (Macgregor *et al.* 2005; Lund *et al.* 2008), models with high order RR polynomials fitted for the QTL effect presented convergence problems. Moreover, the abovementioned simulation studies did not address the need for effective modelling of other time-dependent effects to accurately describe QTL effects over time.

To further develop and evaluate the RR technique for identification and description of age-dependent QTL for the longitudinal trait of growth, an application of RR models on actual phenotypes was pursued and described in chapter 5. Specifically, RR models were fitted to live weights over time from Scottish Blackface sheep to identify QTL. The study in chapter 5 provided inconclusive results for detection of time-dependent QTL on different chromosomes. That study also showed that RR model choice was complex, even without inclusion of a QTL effect. Additionally, dissection of the QTL variance profile over time was not possible for apparently significant QTL, presumably due to inflated variance predictions at the end points of the time range due to prediction using regression (also observed in e.g. Meyer 1998; Fischer *et al.* 2004a,b).

The main objective of the work presented in the current chapter was to further scrutinise issues that were revealed in chapter 5 relating to the use of RR for time-

dependent growth QTL detection. To facilitate this, a more realistic setup of longitudinal trait components than in previous simulation studies was designed. Specifically, simulated growth phenotypes were composed of time-dependent QTL, polygenic, permanent and temporary environmental effects. Four separate scenarios of QTL time-dependence were simulated to assess whether RR analysis would allow QTL detection and description of changes in QTL variance along time. In particular, these simulation scenarios were used to explore further the RR model choice procedure, first introduced in chapter 5. Additionally, the performance of statistically optimal and suboptimal RR models to describe the simulated changes in variance along time for the QTL and other effects was evaluated.

6.2 Materials and Methods

6.2.1 Overall simulation structure

A population of lambs whose growth followed the Gompertz growth curve trajectory (Turner *et al.* 1976; Lambe *et al.* 2006) was simulated, with the genetic control of daily growth rate changing over time. Between-animal genetic variation for growth, superimposed on the Gompertz growth trajectory, consisted of two components: a pseudo-polygenic component, described by variation at ten genes, and a single biallelic QTL effect. The polygenic component was modelled to change linearly over time, such that the genetic correlation between time points decreased with time. The QTL effect was modelled to have various time-dependent trajectories as described below. Further, a permanent animal effect was included, with

correlations decreasing as time between growth measurements increased. Finally, residual variance was added to individual phenotypes at each time point of interest.

6.2.2 Population

The population was simulated to be representative of a commercial sheep pedigree structure and data recording spanned one year. The parental generation was obtained by random sampling without replacement from an unrelated base population of 2000 individuals with equal number of males and females. Sires and dams were chosen at random from the parental generation and dams were randomly allocated to sires. The full dataset contained 16 sires, 60 dams per sire, and 2 offspring per mating, i.e. a total of 1920 lambs.

6.2.3 Simulated genetic and residual effects

Consider initially a trait at a single time point. A chromosome length of 15 cM was simulated. It consisted of 4 markers, each with 4 alleles, spaced at 5 cM intervals and a biallelic QTL with allele frequency of 0.5 and varying effect size, fitted between the 2nd and 3rd marker at 7.5 cM. Ten biallelic loci with additive effects of 0.03, unlinked from each other and from the markers/QTL, were simulated to represent polygenic effects, with allele frequency of 0.5 at each locus, using the method described by Alfonso and Haley (1998). In a strictly polygenic scenario, the phenotypes that were produced were normally distributed, indicating that the ten additive effects sufficiently described a realistic polygenic variance structure. The simulated polygenic variance was 0.0045 (estimated as $2npqa^2$, where n=number of

loci; p,q=allele frequencies; a=additive effect at each locus). A residual effect for each individual, to make up the permanent animal effect subsequently, was sampled from a normal distribution with mean 0 and variance 0.005. No other effects were simulated. An existing Fortran simulation script was used for this stage of the simulation (written by Sara Knott).

6.2.4 Time-dependence structures

To assign time-dependence to the polygenic and non-genetic (permanent animal) effects and to the isolated QTL effect in order to make its expression time-dependent, two simulations were run using different random seeds but the same pedigree structure. In these two simulations, the biallelic additive QTL effect was identified and subtracted from the respective individual value. The resulting values from the two simulations per animal were arbitrarily assigned to be the "start" and "end" phenotypes (sets 1 and 2, respectively), with the "start" assumed to be at day -50 (i.e. prenatal) and "end" at day 250. The marker genotypes from the first simulation were subsequently used for QTL mapping in all analyses.

Time-dependent polygenic and permanent animal effects were created by linear interpolation of sets 1 and 2, which were not correlated. Subsequently (see below) phenotypes were constructed at days 0 (i.e birth), 28, 56, 84, 112, 140, 168, 196, 224, and 252. These time points were chosen to represent monthly measurements from birth and up to 9 months of age for each lamb. The appropriate QTL effect for each time-dependence QTL scenario (see below) was then added to

the interpolated polygenic and permanent animal effect of every animal at each time point to obtain a composite effect f per animal.

Four QTL time-dependence scenarios were studied: a) constant QTL allelic effect equal to 0.06 over time ($QTL_{constant}$); b) QTL effect starting with allelic substitution effect of 0.06 and decreasing linearly over time to 0.0001 (QTL_{high_low}); c) QTL effect starting at 0.0001 and increasing linearly over time to 0.06 (QTL_{low_high}); d) QTL additive effect of 0.0001, increasing linearly and having its highest value of 0.06 at the midpoint of the time range studied and then decreasing linearly to become 0.0001 at the end day ($QTL_{midpoint}$). Graphical representations of the time-dependent trajectories for the QTL effect (within the range of days zero to 252) are shown in Figures 6.1a-6.1d. The patterns of QTL variances over time that arise from these effects are depicted in Figure 6.2. Additionally, expected estimates of the QTL variances at day zero, mid- and end-points for the different scenarios are given in Table 6.1.

Table 6.1 Expected QTL variances over time for the four simulated scenarios

Simulation Scenario[1]	σ^2_{day0}[2]	σ^2_{day100}[3]	σ^2_{day252}[4]	Approximate change in QTL variance
$QTL_{constant}$	0.0018	0.0018	0.0018	None
QTL_{high_low}	0.0013	0.00045	$4.5x10^{-8}$	Linear
QTL_{low_high}	$5x10^{-5}$	0.00045	0.0018	Linear
$QTL_{midpoint}$	0.0002	0.0018	$2.5x10^{-7}$	Triangular

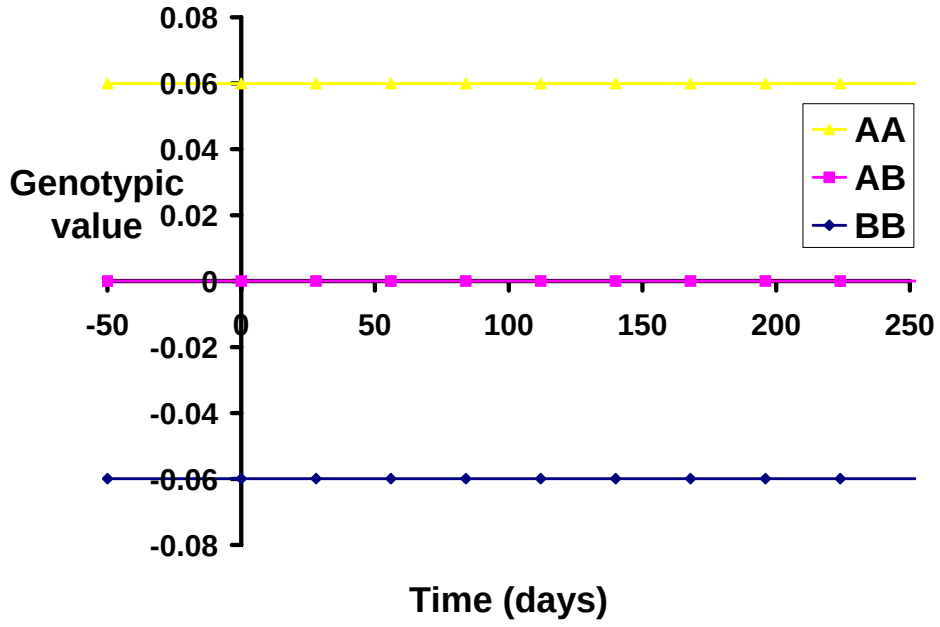
[1]In all scenarios: expected polygenic variance σ^2_α=0.0045; expected permanent animal variance sampled from N(0, 0.005).

[2]QTL variance at day 0.

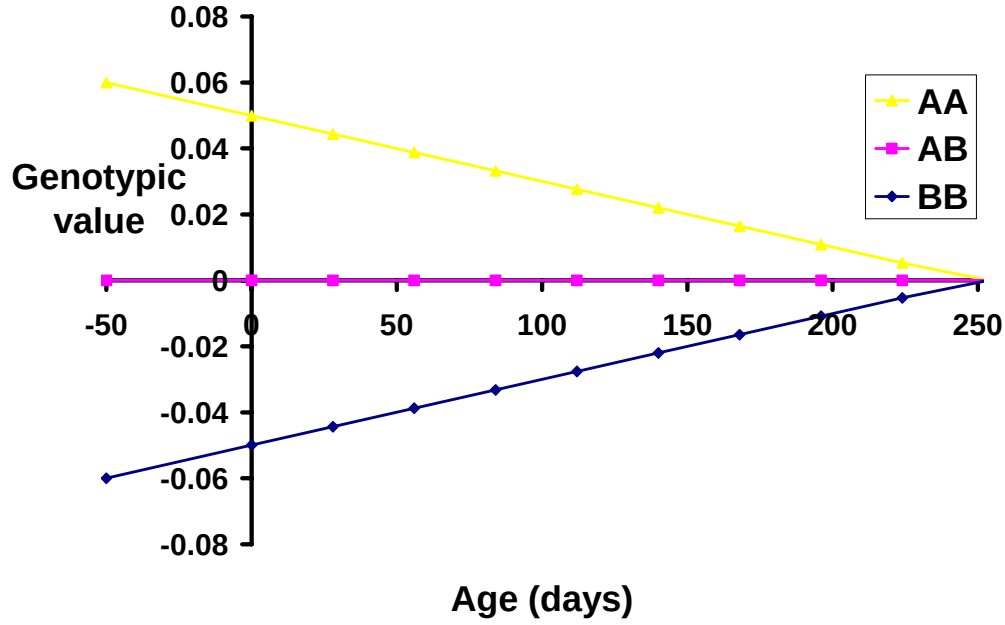
[3]QTL variance at day 100.

[4]QTL variance at day 252.

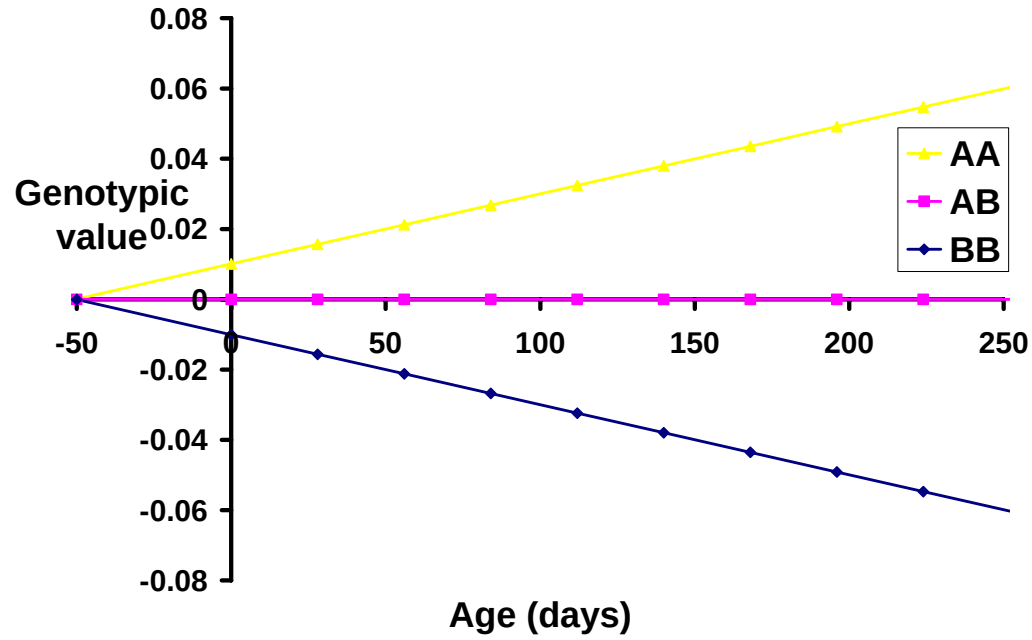
a


0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
Genotypic value
-50
0
50
100
150
200
250
AA
AB
BB
Time (days)

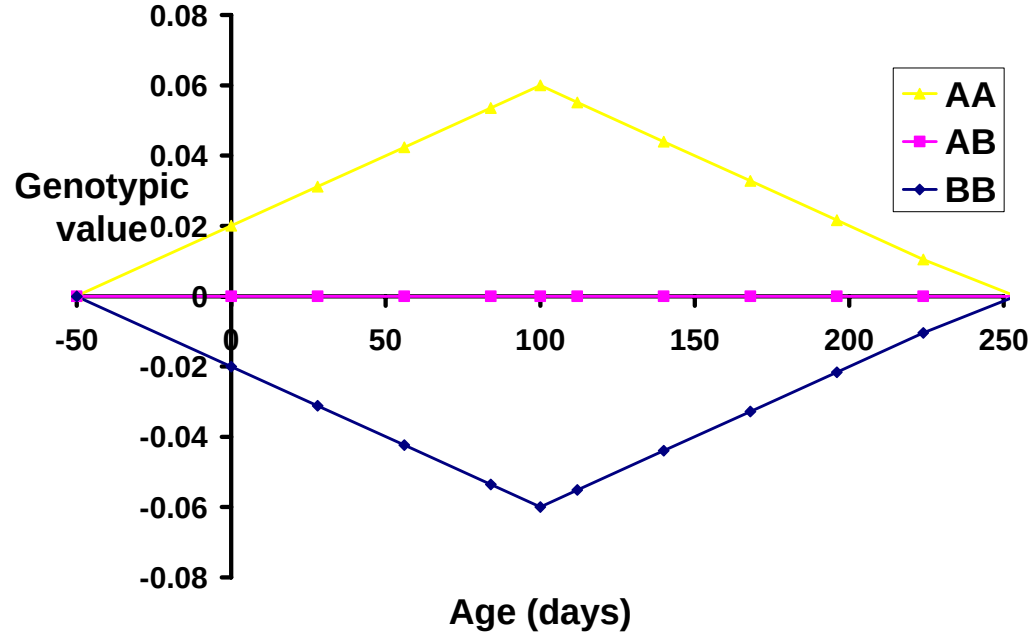
b


0.08
0.06
0.04
0.02
0
-0.02
-0.04
-0.06
-0.08
Genotypic value
-50
0
50
100
150
200
250
AA
AB
BB
Age (days)

c



d



Figures 6.1a-6.1d Change in magnitude of effect over time for each of four QTL time-dependent trajectories simulated: a) $QTL_{constant}$; b) QTL_{high_low}; c) QTL_{low_high}; d) $QTL_{midpoint.}$ AA corresponds to the favourable QTL genotype.

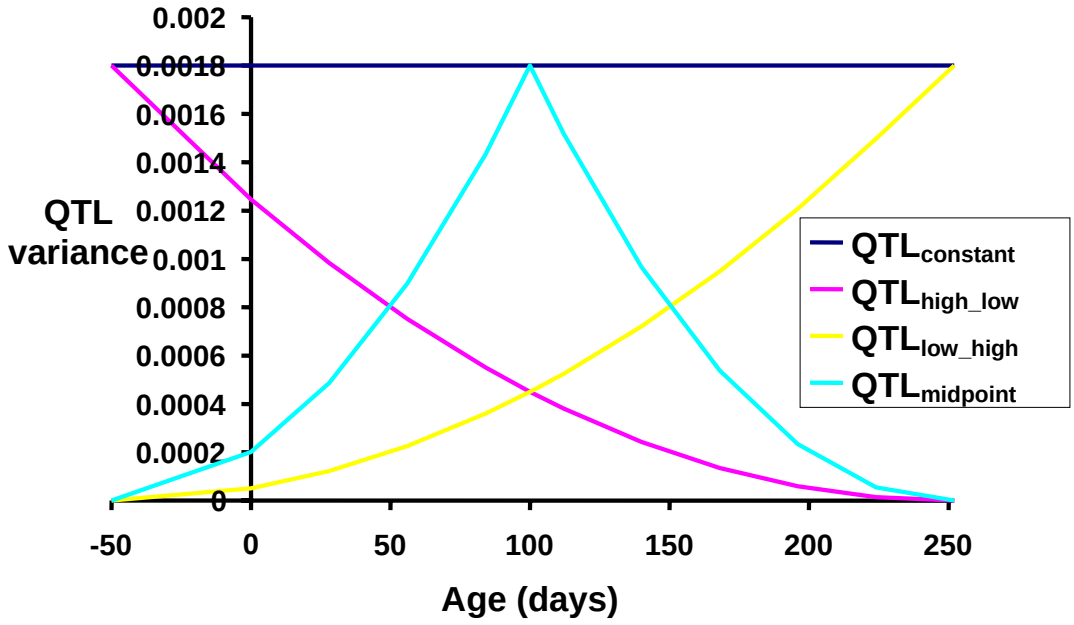


Figure 6.2 Simulated QTL variances over time for the four scenarios of QTL age-dependence.

6.2.5 Gompertz growth curve trajectory

The Gompertz growth curve equation was used to model a) weight and b) daily gain over time using the aforementioned simulated composite effects at each time point for each lamb. The parameterisation of the Gompertz model equation used (explained in Lambe *et al*. 2006) and its derivative are the following:

$$y(t) = Ae^{\{-e^{[Be(C-t)/A]}\}} \text{ and } dy/dt = [(B/A)e]y(t)\ln[A/y(t)]$$

y(t) = live weight at time t; A = estimated final body weight, kg; B = maximum growth rate (daily gain), kg/d; C = age at maximum growth rate, d; dy/dt = daily gain at time t.

Estimates of growth parameters A, B, C, were derived from the estimated means across all animals included in the growth curve QTL study of lamb weight

data described in chapter 3 and Hadjipavlou & Bishop (2009). These estimates are given in Table 3.3.

To create daily gain phenotypes over time, each composite effect f at time t (centred on 1 to make all values positive and relative to the population mean), was multiplied by the corresponding Gompertz daily gain (dy/dt; given above) at that time. Random noise was subsequently added to the phenotype by sampling from $N(0, 0.005mean_t^2)$ at each time point, where $mean_t$ was the population mean daily gain at time t.

To create live weights that followed the Gompertz growth trajectory over time, the following equation was used: $\mathbf{y}_{i(t)} = \sum_{-50}^{t}\left[(1+f_{it})\frac{dy}{dt}\right] + \mathbf{e}_{it}$

Where, $\mathbf{y}_{i(t)}$ is the weight of animal i at time t, f is the composite effect of animal i at each time point (described above), and dy/dt is once again the Gompertz curve daily gain. The start time point is -50 days, with summation up to time t. $\mathbf{e}_{it}$ is the simulated random noise at time t, and was sampled from $N(0, 0.005mean(w)_t^2)$, where $mean(w)_t$ was the population mean live weight at time t. The residual term is assumed to comprise random 'noise' effects that are not cumulated across time, such as animal gut fill (a major component of sheep live weight) and measurement imprecision.

Overall, with the above full simulation structure, polygenic heritability was expected to be ≤0.31, and permanent animal variance ratio ≥0.28 across time for the resulting weight and daily gain phenotypes. Further, the QTL heritability was expected to range from a maximum of 0.11 for time points when the QTL effect was highest (a=0.06), down towards zero as the effect decreased.

6.2.6 Random regression model

A random regression (RR) model was used to fit the simulated longitudinal live weight or daily gain, first, without accounting for a QTL and, subsequently, with the inclusion of a QTL effect. The generalized RR equation used for the full model is a reduced form of the equation described in chapter 5:

$$\mathbf{y= X\beta + Wq + Z_1u + Z_2p + Z_3e}$$

where **y** is a vector of observations taken at several time points for each lamb; **β** is a matrix of age-dependent fixed effects, in this case corresponding to a fixed regression on age, describing the population mean trajectory over time; **X** is the design matrix connecting fixed effects with records; **q**, **u**, **p,** are vectors containing systematic time-dependent deviations from the fixed curve, modelled as random effects, due to allelic effects of the QTL, polygenic, and permanent animal effects, respectively, mapped onto the data by design matrices **W**, $\mathbf{Z_1}$, **and** $\mathbf{Z_2}$; **e** is the vector of residual effects, mapped onto the data by matrix $\mathbf{Z_3}$. Vector **q** is of dimensions $2N_gp_1$, where N_g is the number of animals included in the gametic matrix. Vector **u** is of dimension N_ap_2, N_a being the number of animals in the relationship matrix (i.e. pedigree). The permanent animal vector **p** and the residual vector **e** are of dimensions N_pp_3, and N_pp_4, respectively, where N_p is the number of animals with records. Parameters p_1, p_2, p_3, p_4 correspond to the number of random regression coefficients used to model the associated random effect.

Matrices **W**, $\mathbf{Z_1}$ and $\mathbf{Z_2}$ are design matrices of RR curves. The elements of these matrices are Legendre polynomials of specific order for lamb *i* Φ_i. $\Phi_i=\forall_i(t^*_{ij})$, where $\forall_i$ are coefficients of the chosen Legendre polynomial for lamb *i* used to calculate Φ_i at age *j* (t^*_{ij}). The age values are standardized between -1 and +1. The

random vectors **q**, **u**, **p**, **e**, are assumed to be mutually independent and to follow multivariate normal distributions;

$\mathbf{q}_i|M, c_i \sim MVN(0,\mathbf{K}_{0i}\otimes\mathbf{Q}_i|M,c_i)$, $\mathbf{u} \sim MVN(0,\mathbf{G}_0\otimes\mathbf{A})$, $\mathbf{p} \sim MVN(0, \mathbf{P}_0\otimes\mathbf{I})$ and $\mathbf{e} \sim MVN(0, \mathbf{I}\sigma^2_{ek})$, where σ^2_{ek} is the residual variance of age class k.

In the above specification, $\mathbf{K}_{0i}$, $\mathbf{G}_0$, $\mathbf{P}_0$ are covariance matrices among random regression coefficients, **A** is the additive genetic relationship matrix and $\mathbf{Q}_i|M,c_i$ is the gametic relationship matrix of the allelic effects at the QTL, conditional on marker data (M) and the position (c_i) on the chromosome. $\mathbf{K}_{0i}$, $\mathbf{G}_0$, $\mathbf{P}_0$ were specified to be unstructured general covariance matrices and positive definite. The number of (co)variances among the random regression coefficients were $[p_n(p_n+1)]/2$, where $p_n \times p_n$ are the corresponding dimensions of each of the three covariance matrices**.** Thus, p_n corresponds to order+1 of the random regression polynomial fitted; e.g. fitting a polynomial of 2nd order results in three random regression coefficients (coefficients of order zero, one and two) whose variances and pairwise covariances need to be estimated.

To avoid convergence problems due to model overparameterization, reduction of the number of estimated parameters was pursued, while including all relevant variance components. For this purpose, the longitudinal data were subjected to a natural logarithm transformation prior to analyses which allowed the assumption of homogeneous residual variance across time; i.e. one residual variance across the ten monthly intervals with phenotypes. In all models fitted, a fixed regression Legendre polynomial of 6th order for age was deemed appropriate, by sequentially testing the inclusion of a higher-order term. A fixed regression term is the longitudinal equivalent of the population mean fitted in single time-point analyses.

6.2.7 Model building, statistical testing and variance ratio construction

In general, models were assessed by their likelihoods. However, the Likelihood Ratio Test (LRT) requires nested models; hence a systematic approach is needed to test alternative models, as was also described in chapter 5. For this purpose, prior to fitting a QTL, LRT was used to assess the significance of polygenic and permanent animal effects across nested ("no QTL") RR models with varying order of the random regression polynomials for each effect. The LRT statistic was calculated as twice the difference between the log likelihood of the full and the reduced models, and was assumed to have a chi-squared distribution with degrees of freedom equal to the number of extra parameters estimated in the full model compared to the reduced one.

"No QTL" models with the optimal RR order for polygenic and permanent animal effects were then used to fit a QTL effect. In cases for which the statistically best "no QTL" model did not converge upon including a QTL effect, models with reduced polynomial orders for the polygenic and permanent animal effects were also explored. The QTL effect was also modelled with random polynomial curves to allow for systematic effects of the QTL on the deviation of the animal phenotype from the expected population means over time, and hence allow for changes in the QTL variance with age.

As was first described in chapter 5, formal model testing included LRT at the chromosome position with the highest likelihood for each QTL scenario and trait analysed to assess a) overall QTL significance and b) different orders of the RR

polynomial fitted for the QTL. The test statistic was again taken to have chi-squared distribution. This is conservative in the RR framework, as explained by other researchers (e.g. Meyer 1998; Macgregor *et al.* 2005), since the asymptotic distributions are likely mixtures of chi-squared distributions. Additionally, assuming a naïve chi-squared distribution is generally conservative for a test for a QTL at a single location in the genome (Self & Liang 1987; Stram & Lee 1994; Allison *et al.* 1999; Visscher 2006, Macgregor *et al.* 2005).

All RR mixed model analyses were performed using ASReml (Gilmour *et al.* 2002), as detailed in chapter 5. ASReml calculates the inverse of the polygenic relationship matrix **A** from pedigree information, but requires the inverse of the gametic relationship matrix **Q**. Matrix **Q** was calculated using Loki (Heath 1997) at 1 cM intervals along the simulated chromosome of 15 cM, across all animals in the pedigree. **Q** matrix inversion was obtained using R (R Development Core Team 2007). The gametic relationship matrix was estimated using marker information from both sires and dams. One thousand iterations were specified to allow log likelihood maximisation and model convergence in all analyses. For each RR model that included the QTL effect, the likelihood of the full model was maximised every 1cM for each order of RR for the QTL effect.

For every RR analysis that achieved convergence, ASReml provided solutions for the (co)variances among RR coefficients for each time-dependent RR effect (i.e. estimation of matrices $\mathbf{K_0}$, $\mathbf{G_0}$, $\mathbf{P_0}$) and evaluation of the Legendre polynomials at each age point (i.e. construction of matrices $\mathbf{W}$, $\mathbf{Z_1}$, $\mathbf{Z_2}$), fitted to each effect. Following Kirkpatrick *et al.* (1990), and as detailed in chapter 5, the above were used to estimate the (co)variance matrix for each time-dependent effect

separately. The respective (co)variance matrix values could then be used to estimate the proportion of phenotypic variance attributed to a particular variance component (e.g. polygenic heritability) at different time points over time. In addition, correlations between pairs of ages could be calculated for each effect (see chapter 5 for details).

Due to the extremely computer-intensive nature of the analyses and the requirement for manual model inspection and choice, it was not possible to replicate the QTL scenarios.

6.3 Results

6.3.1 Simulation output

Univariate genetic analyses of the strictly polygenic and permanent animal effect output from the first part of the simulation, obtained with two different seeds, resulted in polygenic heritability of 0.51±0.08 for set 1 and 0.30±0.08 for set 2 for the 1920 lambs studied, merely due to chance. This indicated that with linear interpolations between these two simulation outputs, the polygenic component would have lower variance at the end point (set 2) than at the start (set 1). Conversely, the permanent animal effect, which is derived from the residual component, would have increased variance at the end point. These patterns were indeed observed in analyses of the composite phenotypes of weight and daily gain.

A scatterplot of the interpolated polygenic and permanent animal effects across the ten time points studied (prior to further manipulation; i.e. before adding the QTL or creating Gompertz curve-dependent daily gains or live weights), along with variance ratios for the two components and the coefficient of variation (CV) are summarised in Appendix 6.1. Figure 6.3 shows a scatterplot of the composite live weight phenotypes across time from the $QTL_{constant}$ scenario. Phenotypic variances and polygenic heritabilitites derived from univariate analyses, not accounting for the QTL, at each of the ten time points for this scenario are given in Table 6.2. These were obtained using ASReml (Gilmour *et al.* 2002). The corresponding estimates obtained from analyses of the other three QTL scenarios are given in Appendix 6.2. Genetic correlations over time were estimated for all scenarios using (co)variance estimates from the chosen "no QTL" RR models (see below). In all cases, genetic correlations decreased as the interval between measurements increased; specifically for live weight, these ranged from 0.99 between weight at birth and at 28 days, and zero between the trait at birth and at 252 days (results not shown).

Table 6.2 Phenotypic means, variances and polygenic heritabilities of live weight at the ten time points studied for the $QTL_{constant}$ scenario [1]

Age (days)	Phenotypic mean (kg)	Phenotypic variance (kg^2)	Heritability ($h^2 \pm$ s.e.)
0	3.82	0.20	0.46 ± 0.08
28	10.72	1.43	0.39 ± 0.08
56	18.67	4.35	0.40 ± 0.08
84	25.12	7.38	0.35± 0.08
112	29.55	10.08	0.42 ± 0.09
140	32.21	12.43	0.32 ± 0.08
168	33.75	13.73	0.33 ± 0.08
196	34.58	15.23	0.28 ± 0.07
224	35.02	17.42	0.34 ± 0.08
252	35.25	19.90	0.31 ± 0.08

[1] Coefficient of variation (CV)=$\sigma(t_i)/\mu(t_i)$) was 0.1 for phenotypes at all ages

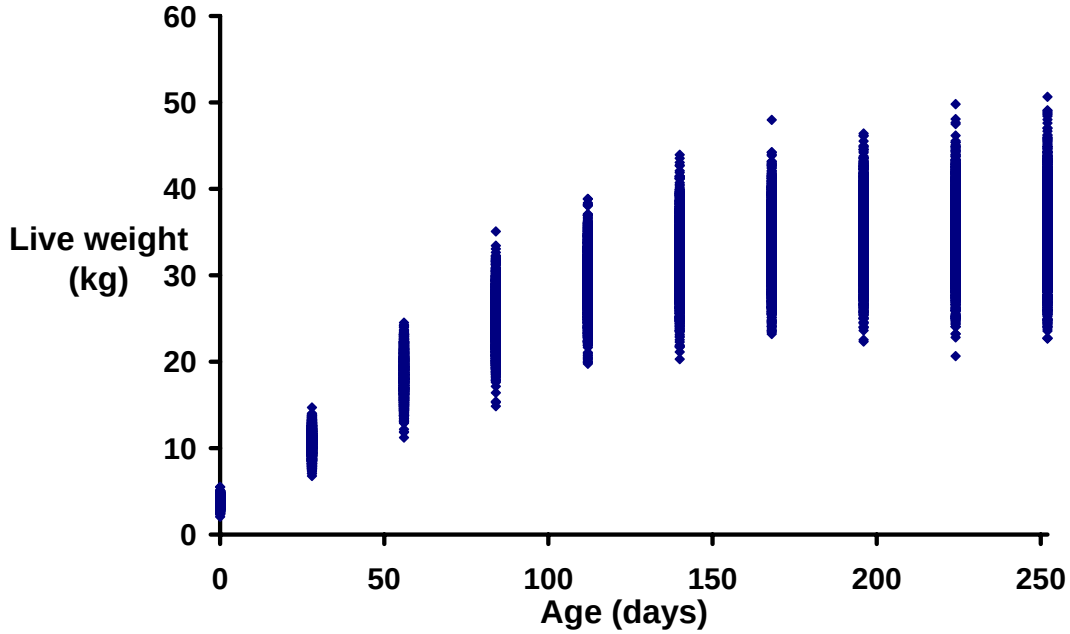


Figure 6.3 Scatterplot of live weights for the 1920 lambs across the ten time points studied for the $QTL_{constant}$ scenario.

6.3.2 Random regression analyses

RR analysis of untransformed phenotypes was not pursued to minimise the possibility of model overparameterisation and convergence issues that seemed to arise in corresponding analyses in chapter 5. Therefore, results reported below refer to RR analyses of transformed data, using natural logarithm transformation, with homogeneous residual variance and an unstructured (US) covariance matrix specified for the random regression coefficients of each time-dependent effect.

Appendix 6.3 contains the LogL estimates for the "no QTL" RR models fitted to weights and daily gains separately, for each of the four different QTL structures simulated over time. Nested LRT could only be performed across rows or down columns, and the degrees of freedom (df) associated with RR orders 0, 1, 2, 3 were

1, 3, 6, 10, respectively. Some high order "no QTL" models did not converge (Appendix 6.3) but the convergence properties for the remaining models followed expectation, with the log likelihood increasing when more parameters were estimated in the models. In cases in which the LogL did not significantly improve when the order of either polygenic or permanent animal effect was increased, the higher order model was rejected.

Table 6.3 Summary of chosen "no QTL RR models" for all scenarios studied

Time-dependence scenario	Trait	Random regression model[1]	No. model parameters[2]	Log likelihood (LogL)[3]
$QTL_{constant}$	Weight	11	7	7223
	Daily gain	11	7	7131
QTL_{high_low}	Weight	11	7	7368
	Daily gain	11	7	7415
QTL_{low_high}	Weight	11	7	7380
	Daily gain	11	7	7183
$QTL_{midpoint}$	Weight	21	10	7029
	Daily gain	21	10	7267

[1]Polynomial order for direct additive and permanent animal effects, respectively.

[2]The model parameters estimated for each scenario correspond to sum of [(n+1)(n+2)]/2 for each random term fitted with RR plus one residual variance; n=polynomial order for each effect.

[3]LogL values presented for comparison with subsequent results.

Table 6.3 provides a summary of the chosen "no QTL" RR model, based on LRT results, for each scenario and each trait examined. A model with a 1st order RR polynomial fitted for both the polygenic and permanent animal effect was deemed statistically optimal for analysing both weight and daily gain in the $QTL_{constant}$, QTL_{high_low}, and QTL_{low_high} scenarios. A model with a 2nd order RR for the polygenic and a 1st order for the permanent animal effect was the optimal "no QTL" RR model for both phenotypes in the $QTL_{midpoint}$ scenario (Table 6.3; Appendix 6.3). Presumably in this case the quadratic polynomial fitted for the polygenic effect variance may be capturing some of variance of the QTL effect.

6.3.3 Performance of full random regression models

Table 6.4 contains a summary of statistically optimal full models, upon inclusion of gametic relationship information by fitting a time-dependent QTL effect in the previously chosen "no QTL" model for each scenario and each phenotype. In each case, models with increasing RR order for the QTL were compared. LRT was used to assess the significance of including a QTL in the corresponding "no QTL" model (Table 6.4) and for increasing the QTL polynomial order (Appendix 6.4).

Subsequently, using the estimated parameters from the full models, the predicted proportions of phenotypic variance explained by each effect at each of the ten simulated monthly intervals were calculated and plotted over time, in order to compare the magnitudes and longitudinal trends with expectations for each scenario.

Table 6.4 Summary of chosen QTL RR models for all scenarios studied

Time-dependence scenario	Trait	RR model[1]	No. model parameters[2]	Max. LogL position (cM)	Highest LogL	LRT
$QTL_{constant}$	Weight	111	10	8	7288	130 (3 d.f.)
	Daily gain	111	10	9	7198	135 (3 d.f.)
QTL_{high_low}	Weight	110	8	8	7381	26 (1 d.f.)
	Daily gain	111	10	6	7442	52 (3 d.f.)
QTL_{low_high}	Weight	111	10	8	7408	56 (3 d.f.)
	Daily gain	111	10	7	7208	50 (3 d.f.)
$QTL_{midpoint}$	Weight	211	13	6	7050	42 (3 d.f.)
	Daily gain	211	13	9	7287	40 (3 d.f.)

[1]Polynomial order for direct additive, permanent animal and QTL effects, respectively.

[2]The model parameters estimated for each scenario correspond to sum of [(n+1)(n+2)]/2 for each random term fitted with RR plus one residual variance; n=polynomial order for each effect.

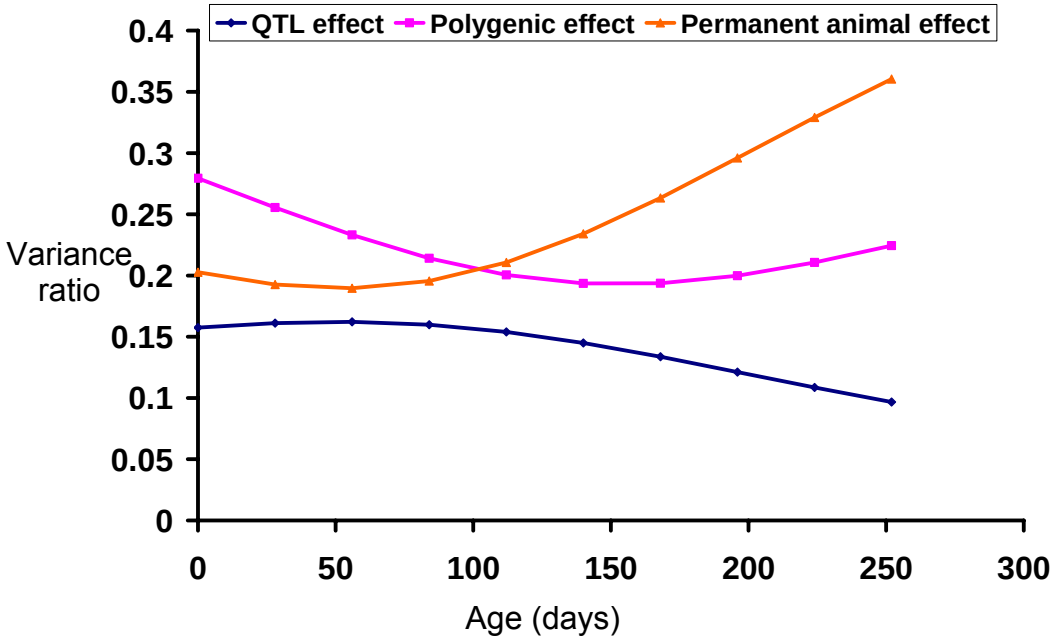
6.3.3.1 QTL with constant effect over time

For $QTL_{constant}$, analysis of weight resulted in a zero order RR for the QTL effect to be highly significant ($P<0.001$) and increasing the QTL RR order to one was borderline significant (LRT=6 (2 d.f.); $P<0.05$). In both models, the highest LogL was obtained at 8 cM, even though in this and the subsequent results, the LogL at the position of highest value was never statistically better than the other positions in the simulated 15cM-interval. For the same scenario, daily gain analysis provided analogous results, with zero order for the QTL effect being highly significant

throughout the chromosome, with highest LogL at 9 cM. Yet, increasing the RR order from zero to one was significant (LRT=7 (2 d.f.); P<0.05). Fitting a 2nd RR order for the QTL led to no convergence for weight and did not improve the fit for daily gain (Appendix 6.3).

The estimated QTL variance trajectory over time was roughly constant in the chosen models, with some variance decrease at higher ages (Figures 6.4a-6.4b). For both phenotypes, the QTL variance trajectory did not differ substantially between a zero and a first order RR for the QTL effect (Figures 6.4a-6.4b; Appendix 6.5). In all models, polygenic heritability decreased over time and the phenotypic variance due to the permanent animal effect increased, as expected from the simulation structure of the two components. Thus, with the statistically optimal models, predicted variance ratio trajectories over time were close to expectation for both weight and daily gain. Estimates of heritabilities were slightly lower than those obtained from univariate analyses (Table 6.2) whereas QTL variance ratios seemed marginally high.

a



b

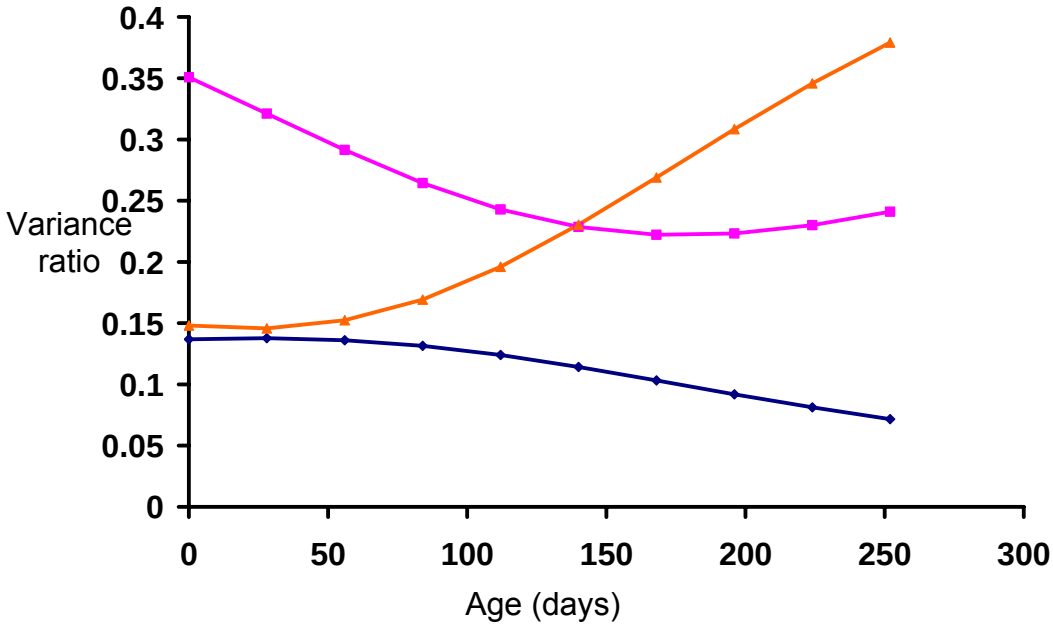


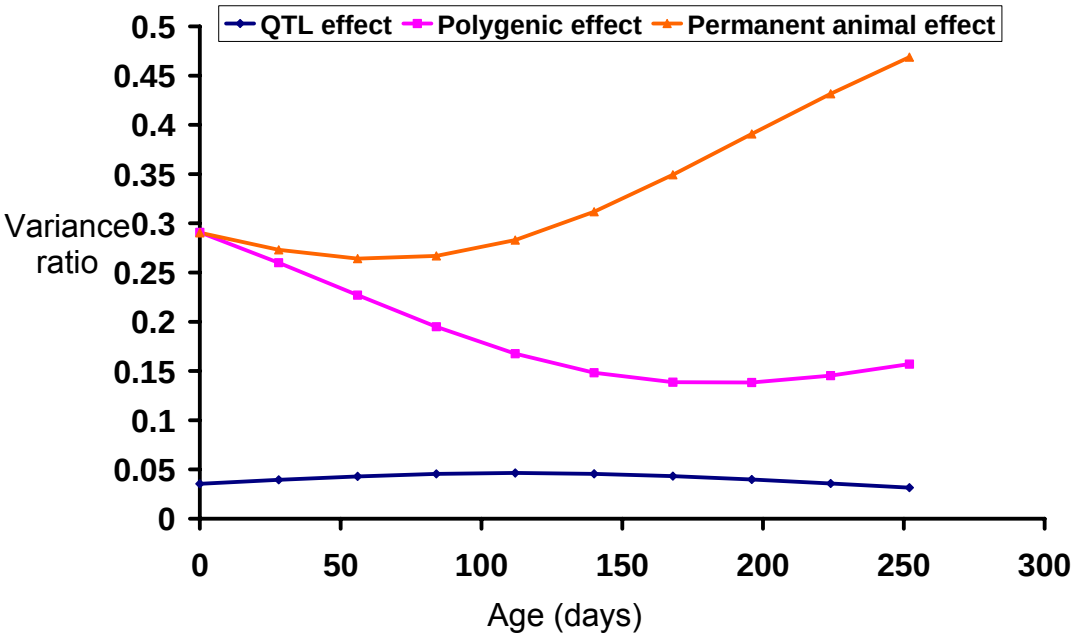
Figure 6.4 Variance ratios over time for effects included in the optimal full models for the $QTL_{constant}$ scenario. Estimates from analyses of a) live weight at 8 cM and b) daily gain at 9 cM. The optimal model included 1st RR order for QTL, polygenic and permanent animal effects for both phenotypes.

6.3.3.2 QTL with decreasing effect over time

For the QTL_{high_low} scenario, a zero RR order for the QTL effect was initially deemed significant for weight and daily gain using LRT (Table 6.4; Appendix 6.4). Increasing the QTL RR order from zero to one led to no convergence for weight, but to a significant improvement in the LogL for daily gain, even though the higher order model only converged up to 6 cM. A RR model with second order for the QTL did not converge for either phenotype (Appendix 6.4).

Based on the statistically optimal full models, the most likely location of the QTL was at 8 cM for weight (LRT=26 (1 d.f.)) and at 6 cM for daily gain (LRT=52 (3 d.f.)). In these models, the variance trajectories over time for all effects fitted followed the respective simulation structure. The variance increased for the permanent animal and decreased for the polygenic effect along time, as shown in Figures 6.5a, b. For weight, the QTL variance ratio appeared to be constant over time (Figure 6.5a), and maximum QTL variance was estimated to be 5% of the phenotypic variance, i.e. about half the maximum simulated QTL heritability. For daily gain, the variance decreased linearly for the QTL effect, in agreement with expectation based on the simulation setup (Figure 6.5b). QTL heritability over time had an analogous trend and magnitude to that seen for weight, when a zero RR order was fitted for the QTL effect (Appendix 6.5).

a



b

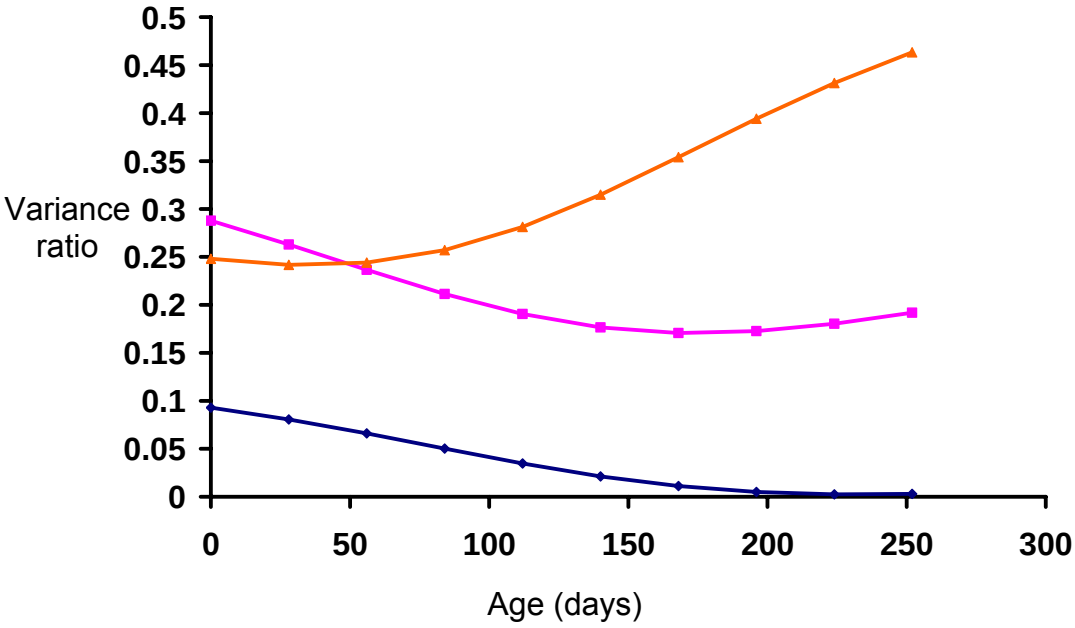
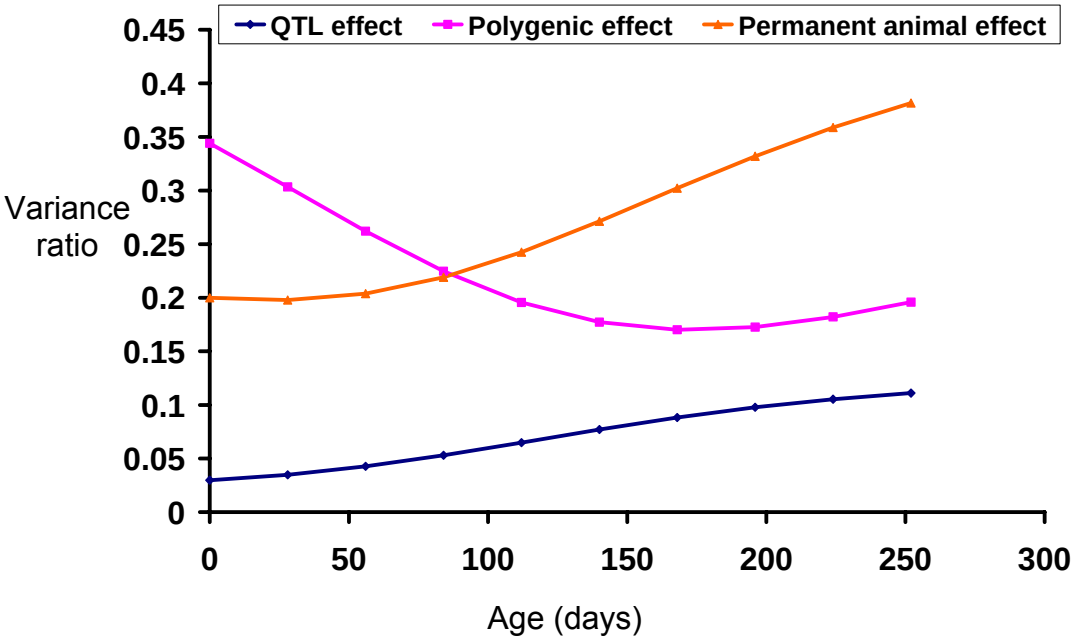


Figure 6.5 Variance ratios over time for effects included in the optimal full models for the QTL_{high_low} scenario. Estimates from analyses of a) live weight at 8 cM and b) daily gain at 6 cM. The optimal model included 1st RR order for polygenic and permanent animal effects for both phenotypes, and zero or 1st order for the QTL for weight and daily gain, respectively.

6.3.3.3 QTL with increasing effect over time

For the QTL_{low_high} scenario, inclusion of a first order RR term for the QTL in the optimal "no QTL" model led to convergence and proved to be statistically better than a zero order term for both weight and daily gain. For weight, the most likely location of the QTL was at 8 cM and at 7 cM for daily gain (Table 6.4). A model with 2nd RR order for the QTL converged for weight, but this model was not significantly better than the one with 1st order for the QTL effect (Appendix 6.4). The same model did not converge for daily gain. The predicted variance trajectories over time fitted expectation based on the simulation scenario for each phenotype, since QTL variance increased approximately linearly with time (Figure 6.6a, b).

a



b

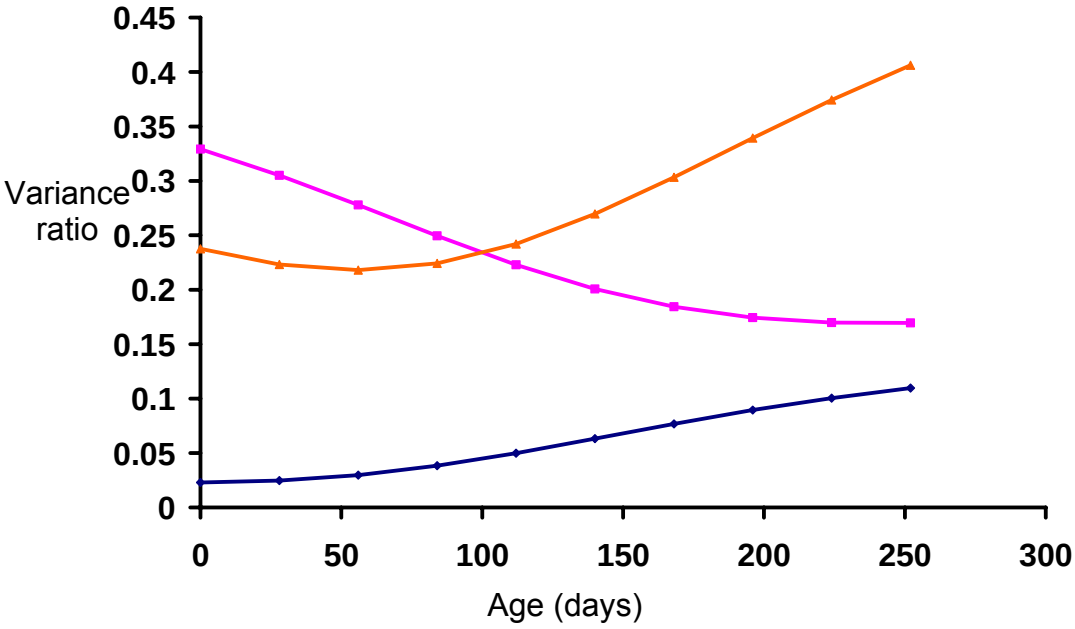


Figure 6.6 Variance ratios over time for effects included in the optimal full models for the QTL_{low_hjgh} scenario. Estimates from analyses of a) live weight at 8 cM and b) daily gain at 7 cM. The optimal model included 1st RR order for QTL, polygenic and permanent animal effects for both phenotypes.

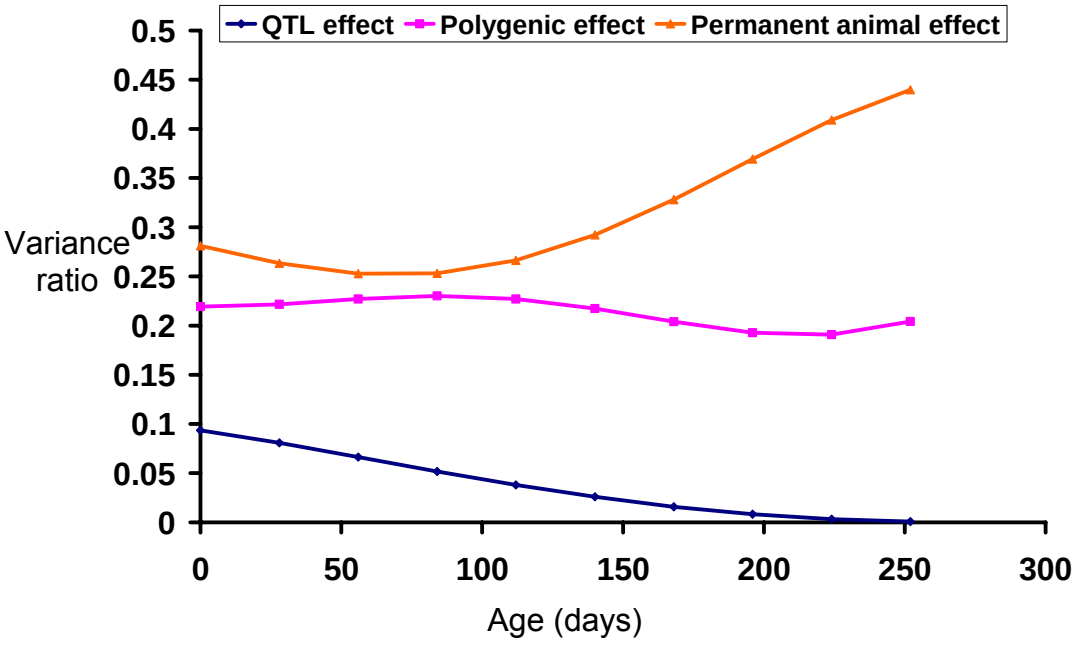
6.3.3.4 QTL with maximum effect at the midpoint of the time range

For the $QTL_{midpoint}$ scenario, as stated earlier, a "no QTL" RR model with 2nd order polynomial for the polygenic effect and 1st order for the permanent animal effect was chosen based on LRT. For the analysis of weight, both a zero and a first order RR polynomial for the QTL resulted in models supporting a highly significant QTL effect, with the higher order model being statistically better (Appendix 6.4). With this model, the most likely position for the QTL was at 6 cM (Table 6.4), although the model did not converge at subsequent positions. The variance trajectory over time for permanent animal effect agreed with the simulated trajectory but did

not follow expectation of near-linear change over time for the polygenic effect (Figure 6.6a). The QTL variance pattern predicted from this model appeared to decrease over time, and became close to zero at 252 days. A second order RR QTL model did not converge, even in the case when the polygenic effect order was reduced to 1 (Appendix 6.3).

When daily gain was analysed, up to a first order RR for the QTL effect resulted in model convergence when the QTL was included in the optimal "no QTL" model and in a model with 1st order for both polygenic and permanent animal effects (Appendix 6.4). A 1st order for the QTL was deemed highly significant in the statistically optimal model and the QTL was positioned at 9 cM (Table 6.4). With this model, the QTL variance trajectory was not representative of daily gain expectation based on the simulation structure, as the estimated QTL variance decreased with time. Fitting a 2nd RR order for the polygenic effect resulted in this effect having its maximum variance value at 100 days. This again did not agree with the simulated polygenic heritability over time, and thus indicated that for this scenario, QTL and polygenic effects were likely confounded. Thus, possibly as a consequence of confounding, the optimal RR model structure did not adequately partition the QTL from the polygenic variance, and the quadratic polygenic effect captured some of the variance of the maximum QTL effect mid growth. For this scenario, confounding potentially persisted in statistically suboptimal models as well, since models with 2nd order for the QTL and 1st or lower polygenic RR order did not converge.

a



b

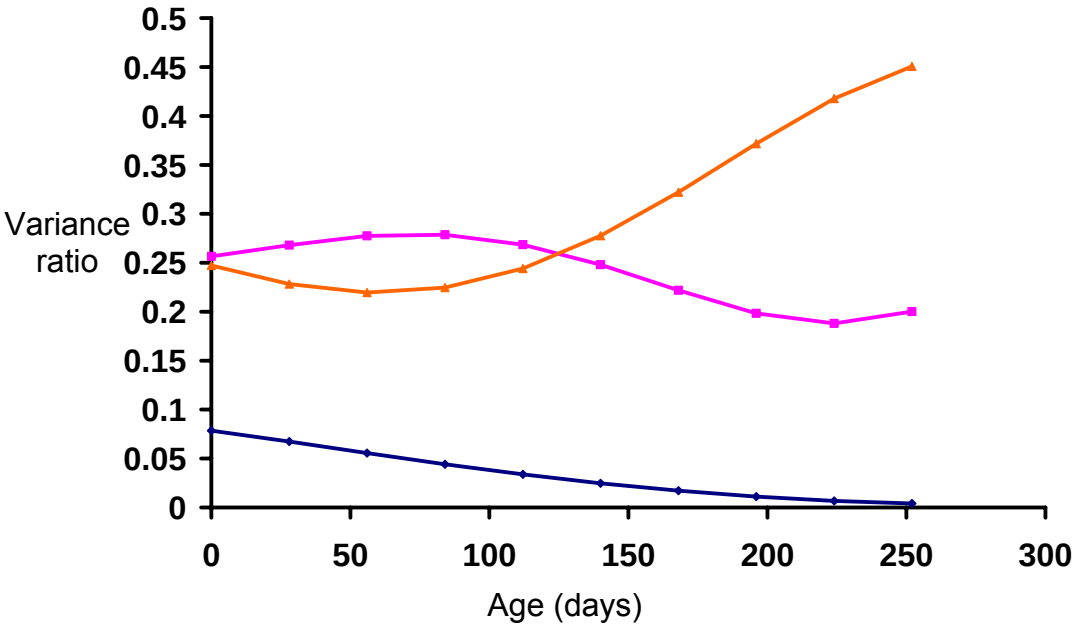


Figure 6.6 Variance ratios over time for effects included in the optimal full models for the $QTL_{midpoint}$ scenario. Estimates from analyses of a) live weight at 6 cM and b) daily gain at 9 cM. The optimal model included 1st RR order for QTL and permanent animal effects and 2nd order for polygenic effect for both phenotypes.

6.4 Discussion

Four simulation scenarios that differed in the pattern of QTL time-dependence for growth were analysed using RR techniques to gain more insight into the ability of RR to dissect QTL whose expression changed over time. The four scenarios were simple cases of what presumably happens in reality, since only a single QTL with time-varying effect on a trait was examined in each scenario and smooth, mainly linear, changes over time, in terms of the QTL effects on daily gain, were simulated. In the most complex structure, the favourable QTL genotype had its maximum effect at the midpoint of the time range studied, analogous to growth curve QTL results obtained in chapter 3. Complementation of the longitudinal QTL effect with time-dependent polygenic and environmental effects, and construction of weight and growth rate phenotypes that followed the Gompertz curve trajectory, resulted in more realistic biological scenarios than previous simulations (Lund *et al.* 2002; Macgregor *et al.* 2005).

To examine the performance of RR models in detecting QTL time-dependence in these scenarios, a two-stage systematic procedure for model choice, first introduced in chapter 5, was followed. This systematic strategy was pursued to examine whether it can provide guidance as to which "no QTL" models should be considered at the outset; further, whether it can serve as a basis for model evaluation. With this procedure, no prior assumptions about the RR polynomial order to be fitted for an effect were made, and the chosen order was solely based on likelihoods and statistical testing of nested models. In the first step of model selection, "no QTL" models were identified in which statistically optimal RR orders for polygenic and permanent animal effect were included for each scenario and phenotype. Since linear

changes over time were simulated for the polygenic and permanent animal effects in all scenarios, the fact that a 1st order RR polynomial was deemed statistically optimal for the three scenarios in which the QTL effect was constant ($QTL_{constant}$) or changed linearly over time (QTL_{high_low}; QTL_{low_high}) indicated that, for these scenarios, the statistical choice coincided with expectation for the description of variances over time for the two effects. General agreement was also observed between simulated and model predicted variance values at different times for polygenic and permanent animal effects.

For the $QTL_{midpoint}$ scenario, however, a 2nd RR order was deemed optimal for the polygenic effect in the "no QTL" model, even though this effect had been simulated to change linearly over time, similarly to the other scenarios. Thus, even at the first step of model choice, there was an indication that the polygenic effect had captured some of the QTL variance, at least prior to fitting the QTL effect in the RR model.

In the second step of the model selection process, initial focus was on assessing whether overall QTL significance would be revealed upon inclusion of the QTL effect in the chosen "no QTL" models. Irrespective of QTL time-dependence scenario, the full models indeed succeeded in identifying the QTL as significant when this effect was fitted with zero or first RR order in models describing either daily gain or live weight. A zero RR order led to statistically significant QTL, even in cases when a first order for this effect resulted in significant improvement of the model LogL. In contrast, lower QTL orders did not provide evidence for a significant QTL in RR analyses of untransformed live weights over time for Scottish Blackface sheep (chapter 5) or longitudinal milk yield in dairy cattle (Lund *et al.* 2008) in some

cases in which a higher order for the QTL resulted in the effect being significant. A possible explanation for this discrepancy may reside on differences in the underlying biology among the different studies. Mainly, the simulated changes in QTL expression were gradual and long lasting in the current study, whereas more abrupt changes in the expression of the QTL may have been present in the aforementioned analyses of live weight and milk yield. Therefore, gradual, smooth variance changes could be modelled effectively using lower order polynomial, while more rapid and complex time-dependencies required higher RR orders. Further studies need to be conducted in order to explore this hypothesis.

In addition to assessing QTL significance, LRT of nested full models with varying order for the QTL was used in the second step of model selection. This was pursued to investigate whether the optimal "no QTL" models could accommodate RR orders for the QTL that would accurately describe the simulated pattern of the QTL variance trajectory over time for live weight and daily gain. To evaluate the performance of selected RR models in describing the QTL variance over time, it is necessary to identify the expected RR orders for the QTL that would successfully model the deviations from the average growth trajectory of the population due to QTL genotype for each simulated scenario and each phenotype.

For daily gain, the expectations are simpler than for live weight. Short-term changes in QTL variance (i.e. occurring within a day) would affect the composite phenotype for each individual on a particular day. Thus, deviations from the average population growth rate, to be detected using RR along the time range studied, would directly mirror the simulated QTL variance pattern in each case. Specifically, for all scenarios apart from $QTL_{midpoint}$, up to a 1^{st} order RR would be expected to be

optimal for the QTL effect in order to accurately describe the simulated linear deviations from the average curve for daily gain over time. For $QTL_{midpoint}$, a RR order that would allow greater deviations to be detected at intermediate ages than at other time points would be expected to be deemed optimal in daily gain analysis. Specifically, at least a 2^{nd} order would seem plausible, as it would allow modelling of curvature for the deviations of the QTL effect.

With respect to live weight analysis, predicting the expected RR order that would accurately describe the QTL variance trajectory in each scenario is complicated by the fact that live weight at a particular age is the integral of all growth rates that occurred up to that time point. Thus the expected deviations due to QTL genotype from the population sigmoidal curve for live weight will follow the combined changes in QTL variance up to each time point studied. In this respect, a constant QTL effect on daily gain will have increasing impact on weight at each time point. As animals with either the favourable or unfavourable QTL genotype continue to grow, their weights will deviate even more from the average weight curve (i.e. average phenotype of animals that lack the QTL). Thus, potentially a RR order greater than zero could be suitable to model the QTL effect in the $QTL_{constant}$ scenario. For the QTL_{high_low} case, with the QTL effect decreasing along time, the weight deviations of animals that inherited the specific QTL effect will remain roughly constant along the growth trajectory. The reason for this is that, especially after maximum growth rate (which was simulated to occur at about 37 days), no major differences are expected for live weight at adjacent time points, since daily gain is affected by a decreasing QTL effect. Therefore, a zero order for the QTL may possibly be sufficient in this scenario. When it comes to the QTL_{low_high} scenario, the

QTL has an increasing effect on daily gain as time passes. Thus, higher deviations are expected for live weight at later than earlier ages and (at least) a first order for the QTL effect would be required to accurately describe this pattern. For the $QTL_{midpoint}$ scenario, increasing deviations are expected for the animals with the QTL effect up to the point at which the QTL has its maximum effect (i.e. at 100 days). From that point onward, the QTL effect decreases along time and weight deviations at adjacent time points will be comparable as they depend on the sum of daily gain deviations up to that time point. Consequently, up to a 1^{st} RR order can be assumed to accurately fit live weight over time for this scenario. Therefore, specific QTL effects on daily gain need not lead to equivalent effects on live weight. This observation is supported by the results described in chapter 3, even though a different method was used to assess QTL time-dependence. Specifically, in chapter 3, certain age-dependent QTL appeared to be significant at earlier time ranges than the equivalent ones for live weight.

In all cases apart from the analysis of the $QTL_{midpoint}$ scenario for daily gain, the full model with the expected RR order for the QTL effect converged and was deemed to be statistically optimal. Optimal full models that identified significant QTL converged for $QTL_{constant}$, QTL_{high_low} and QTL_{low_high} scenarios for both phenotypes. However, for daily gain analysis in the QTL_{high_low} scenario, the optimal full model converged only up to 6cM, whereas a model with zero QTL order led to convergence at all positions. For the $QTL_{midpoint}$ case, only a 1^{st}, and not a 2^{nd}, RR order for the QTL effect resulted in convergence for both the optimal and reduced "no QTL" models in analysis of daily gain. For the same scenario, analysis of weight using the optimal full model (i.e. 1^{st} RR order for the QTL) led to convergence only

up to 6cM. With respect to the $QTL_{midpoint}$ scenario, as it was mentioned earlier, the statistically optimal "no QTL" model, had a 2^{nd} RR order for the polygenic effect, indicating some confounding with the QTL effect. It is likely that apparent confounding may be preventing accurate description of the QTL variance over time for daily gain, and impeding convergence at all positions for the optimal full model for weight.

Thus, even though biologically suboptimal models seemed to nonetheless result in QTL significance in all scenarios, the actual pattern of QTL variance over time appeared to have only been described by RR models that accurately predicted the variance trajectories for the other variance components simulated (polygenic and permanent animal effect) in addition to the QTL. Based on this observation, it can be proposed that in order for RR analysis to account for the covariance structure among measurements ordered in time (i) to identify significant QTL and (ii) also to provide reliable predictions of the QTL variance trajectory over time, it is imperative that other time-dependent components are accounted for in the full models. However, as was shown in chapter 5, optimal RR model choice with no prior information on variance component time-dependence remains complex for field data. In this situation, the two step procedure for RR model choice proposed in chapter 5 and further examined in this chapter could still serve as a starting point.

Confounding, in terms of inefficient partitioning among genetic and non-genetic variance components, is conceptually not surprising and has also been observed in chapter 5 and proposed in other studies (e.g. Lewis & Brotherstone 2002; Fischer *et al.* 2004b). Partial confounding between the QTL and permanent animal effects may be expected in this population structure as both are correlated with an

individual's Mendelian sampling term. In addition, when QTL inheritance is not adequately reconstructed or accounted for, the gametic relationship matrix may be reverting toward the additive genetic relationship matrix, which describes the inheritance of the polygenic effect. In this case, the QTL would then be strongly confounded with the polygenic effect. In RR, the consequences of confounding may be more prominent than in variance component models fitting stationary QTL and other effects. This is because the RR order fitted for an effect partially confounded with the QTL will affect the polynomial order that can be used to model the QTL itself. Thus, due to confounding, even if the QTL is deemed significant, its RR order may be lower than the one that would accurately describe the QTL variance trajectory over time.

Adequate model choice and confounding between effects are not the sole factors affecting convergence and selection of the RR model that would accurately describe the time-dependent random effects. Choice of statistically optimal RR models also depends heavily on the quantity and structure of phenotypes (e.g. Macgregor et al. 2005; Molina et al. 2007) and on marker information. Differences in data availability may in part be the reason that RR analyses of simulated scenarios for daily gain and live weight were seemingly more successful than the Blackface live weight study (chapter 5), since more phenotypes (19200 vs. ~4000) and marker information from both parents were used in the simulations than in the Blackface weight analyses. To examine these impacts, a reduced dataset that included 9 half-sib families to resemble more closely the Blackface data structure, i.e. information on 1080 lambs, but marker data from both parents, was analysed. In all scenarios studied, the same "no QTL" models were deemed statistically optimal as the ones

obtained for the analyses of the full simulated dataset. Overall, analogous results to those from full phenotypic analyses were obtained for including different orders of the QTL effect in the optimal "no QTL" models for RR analysis of reduced phenotypic information. In another exploratory analysis, information reduction was achieved by estimating the gametic relationship matrix using marker information only from the sires, whereas lamb phenotypes from all 16 sire families were studied. On the whole, reduction of the IBD information did not have a detrimental effect in full model convergence and QTL significance for all scenarios. However, it must be noted that only four fully informative markers, each with four alleles, were used in the simulation study, whereas microsatellite markers with variable allele number, information content and varying density on each chromosome were available to map QTL for longitudinal live weights of Scottish Blackface sheep.

In addition to the above, the marked variance inflation that was observed at the edges of the time range studied in chapter 5 was not as evident in the current study, in particular for the QTL variance trajectories. This improvement might in part have risen from the fact that equal phenotypic information was simulated at all time points and that transformed phenotypes were analysed. Moreover, only (up to) linear RR polynomials for the simulated QTL effect in each scenario led to convergence. In contrast, quadratic or higher order polynomials fitted variance components in untransformed weight analyses in chapter 5. High order polynomials have been previously found to provide erratic variance component estimations at later time points, especially when few records were available at extreme ages (e.g. Fischer & van der Werf 2002; Meyer & Kirkpatrick 2005).

Overall, analyses of simulated data provided more comprehensible results than the study of Blackface live weights presented in chapter 5, with respect to dissecting QTL significance and predicting changes of variance components over time. However, the complexity in choosing the biologically and statistically optimal RR model for time-dependent QTL dissection was again emphasized by the simulation work.

6.5 Conclusions

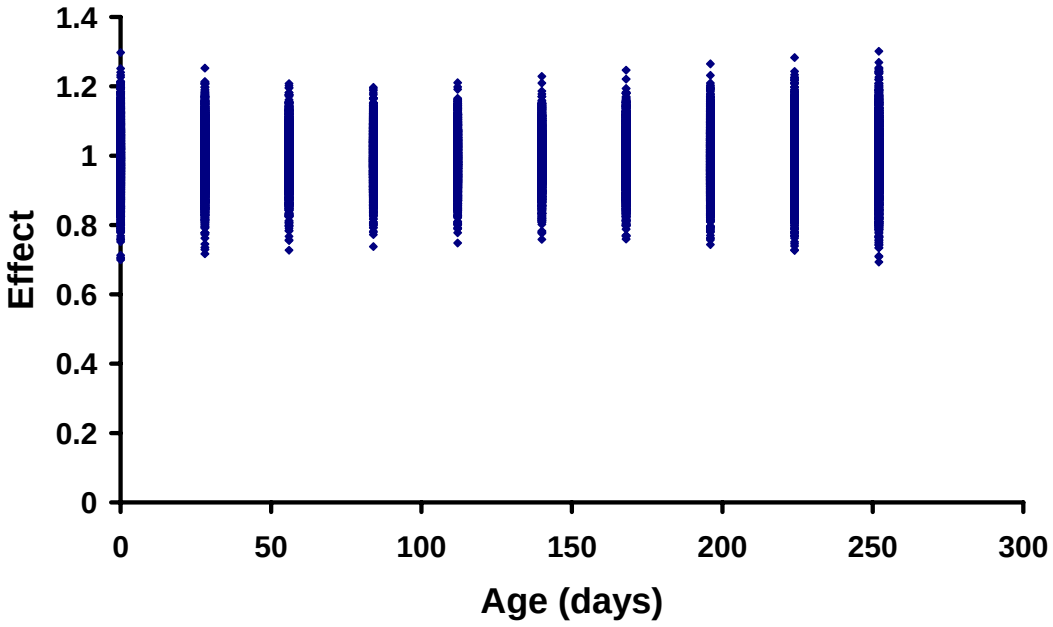
The simulation study provided further insight on the performance of random regression methods for detection of time-dependent growth QTL. Irrespective of QTL time-dependence scenario, full models succeeded in identifying the QTL as significant when this effect was fitted with low RR polynomial orders in either daily gain or live weight analyses. In all cases apart from the scenario where the QTL had maximum effect mid growth, full models with QTL RR order that effectively described the pattern of QTL variance over time converged and were deemed to be statistically better than reduced models. Additionally, in the chosen models, predictions of variance estimates and trajectories generally coincided with expectations for polygenic and permanent animal effects.

Taken altogether, the results of this study emphasised the importance of judicious modelling in the RR framework and, in particular, of accounting for other relevant time-dependent variance components in full models. Only then RR models allowed both the detection of significant QTL and reliable predictions of the QTL variance trajectory over time. However, even in the limited number of simulation

scenarios explored, QTL variance predictions followed the expected pattern only when the change in QTL expression was straightforward, i.e. gradual and roughly linear over time, and no confounding with the other time-dependent effects was apparent. Unless these conditions can be assumed *a priori* in practice, it is difficult to rely on the characterisation of age-dependent QTL solely from statistically optimal RR models.

Appendix 6.1

Graphical representation and variance component estimations of composite polygenic and permanent animal effect interpolations for the 1920 lambs across the ten time points studied. The composite effects at each time point are scaled to have a mean of 1.



Age (days)	Phenotypic variance (Vp)	Coefficient of Variation (CV)[1]	Heritability ($h^2 \pm$ s.e.)
-50	0.01	0.10	0.51 ± 0.08
0	0.007	0.09	0.49 ± 0.08
28	0.006	0.08	0.48 ± 0.08
56	0.005	0.07	0.45 ± 0.08
84	0.005	0.07	0.42 ± 0.08
112	0.005	0.07	0.38 ± 0.08
140	0.005	0.07	0.35 ± 0.08
168	0.005	0.08	0.32 ± 0.08
196	0.006	0.08	0.30 ± 0.08
224	0.008	0.09	0.30 ± 0.08
252	0.009	0.10	0.30 ± 0.08

[1]Coefficient of variation (CV)=$\sigma(t_i)/\mu(t_i)$

Appendix 6.2

Total phenotypic variance and polygenic heritability for live weight in the QTL_{high_low}, QTL_{low_high}, and $QTL_{midpoint}$ scenarios, obtained from univariate analyses at each time point. The coefficient of variation was 0.1 for phenotypes at all ages.

QTL_{high_low} scenario

Age (days)	Phenotypic variance (Vp)	Heritability ($h^2 \pm$ s.e.)
0	0.20	0.33 ± 0.07
28	1.31	0.30 ± 0.07
56	3.87	0.39 ± 0.08
84	6.54	0.26 ± 0.07
112	8.84	0.23 ± 0.06
140	9.97	0.20 ± 0.06
168	11.73	0.16 ± 0.06
196	13.64	0.19 ± 0.07
224	15.46	0.18 ± 0.06
252	17.96	0.15 ± 0.06

QTL_{low_high} scenario

Age (days)	Phenotypic variance (Vp)	Heritability ($h^2 \pm$ s.e.)
0	0.18	0.39 ± 0.08
28	1.24	0.36 ± 0.08
56	3.36	0.25 ± 0.07
84	6.38	0.30 ± 0.08
112	8.84	0.29 ± 0.08
140	11.24	0.27 ± 0.08
168	12.74	0.27 ± 0.07
196	15.47	0.36 ± 0.08
224	17.32	0.26 ± 0.07
252	19.65	0.25 ± 0.07

$QTL_{midpoint}$ scenario

Age (days)	Phenotypic variance (Vp)	Heritability ($h^2 \pm$ s.e.)
0	0.18	0.33 ± 0.08
28	1.28	0.31 ± 0.07
56	3.91	0.34 ± 0.07
84	7.43	0.38 ± 0.08
112	9.83	0.40 ± 0.08
140	11.22	0.31 ± 0.08
168	12.54	0.27 ± 0.08
196	14.00	0.17 ± 0.06
224	16.01	0.14 ± 0.05
252	17.93	0.25 ± 0.08

Appendix 6.3

Log-likelihood estimates for "no QTL" RR models fitting live weight or daily gain over time. Log-likelihood estimates for the statistically optimal model in each occasion are shown in bold.

$QTL_{constant}$ scenario-Weight analysis

Random regression polynomial order[1]		Permanent animal effect			
		None	0	1	2
Polygenic effect	0	6101	6113	7133	NC[2]
	1	7195	7208	**7223**	NC
	2	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

$QTL_{constant}$ scenario-Daily gain analysis

Random regression polynomial order[1]		Permanent animal effect				
		None	0	1	2	3
Polygenic effect	0	5945	5958	7041	7043	NC[2]
	1	7105	7116	**7131**	7132	NC
	2	7106	7117	NC	NC	NC
	3	NC	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

QTL_{high_low} scenario-Weight analysis

Random regression polynomial order[1]				Permanent animal effect	
		None	0	1	2
Polygenic effect	0	6020	6043	7294	NC[2]
	1	7324	7349	**7368**	NC
	2	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

QTL_{high_low} scenario-Daily gain analysis

Random regression polynomial order[1]				Permanent animal effect		
		None	0	1	2	3
Polygenic effect	0	6026	6048	7328	7326	7329
	1	7377	7400	**7415**	7416	7420
	2	NC[2]	NC	7416	NC	NC
	3	NC	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

QTL_{low_high} scenario-Weight analysis

Random regression polynomial order[1]				Permanent animal effect	
		None	0	1	2
Polygenic effect	0	6092	6111	7290	7291
	1	7350	7368	**7380**	7380
	2	7350	7368	7380	7381

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

QTL_{low_high} scenario-Daily gain analysis

Random regression polynomial order[1]				Permanent animal effect	
		None	0	1	2
Polygenic effect	0	5906	5923	7101	7102
	1	7148	7165	**7183**	7185
	2	7149	7165	7183	NC[2]

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

$QTL_{midpoint}$ scenario-Weight analysis

Random regression polynomial order[1]				Permanent animal effect	
		None	0	1	2
Polygenic effect	0	5797	5817	6884	6908
	1	6937	6958	6973	6994
	2	6992	7015	**7029**	NC[2]

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

$QTL_{midpoint}$ scenario-Daily gain analysis

Random regression polynomial order[1]				Permanent animal effect		
		None	0	1	2	3
Polygenic effect	0	6065	6082	7115	7158	NC[2]
	1	7159	7176	7193	7231	7238
	2	7231	7250	**7267**	NC	NC
	3	NC	NC	NC	NC	NC

[1]RR polynomial order is equivalent to order+1 random coefficients estimated.

[2]NC=Model fit did not converge.

Appendix 6.4

Log-likelihood estimates for RR models including a QTL effect for longitudinal live weight or daily gain analyses in the four QTL scenarios studied. RR polynomial order is equivalent to order+1 random coefficients estimated. Estimates of maximum log-likelihood and position of maximum log-Likelihood are given for each QTL RR order.

$QTL_{constant}$ scenario-Weight analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	11	0 7285 (8cM)	1 7288 (8cM)	2 NC[1]

[1]NC: Model fit did not converge.

$QTL_{constant}$ scenario-Daily gain analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	11	0 7195 (9cM)	1 7198 (9cM)	2 7198 (9cM)

QTL_{high_low} scenario-Weight analysis

Random regression polynomial order		QTL effect	
Polygenic and permanent animal effect	11	0 7381 (8cM)	1 NC[1]

[1]NC: Model fit did not converge.

QTL_{high_low} scenario-Daily gain analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	11	0 7426 (14-15cM)	1 7442[1] (6cM)	2 NC[2]

[1]Model LogL converged up to 6cM.

[2]NC: Model fit did not converge.

QTL_{low_high} scenario-Weight analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	11	0 7397 (8cM)	1 7408 (8cM)	2 7411 (7cM)

QTL_{low_high} scenario-Daily gain analysis

Random regression polynomial order		QTL effect	
Polygenic and permanent animal effect	11	0 7195 (7cM)	1 7208 (7cM)

$QTL_{midpoint}$ scenario-Weight analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	21	0 7040 (14-15cM)	1 7050[1] (6cM)	2 NC[2]
	11	6988 (14-15cM)	6998[1] (6cM)	NC

[1]Model LogL converged up to 6cM.

[2]NC: Model fit did not converge.

$QTL_{midpoint}$ scenario-Daily gain analysis

Random regression polynomial order		QTL effect		
Polygenic and permanent animal effect	21	0 5182 (8cM)	1 7287 (9cM)	2 NC[1]
	11	7211 (14-15cM)	7222 (9cM)	NC

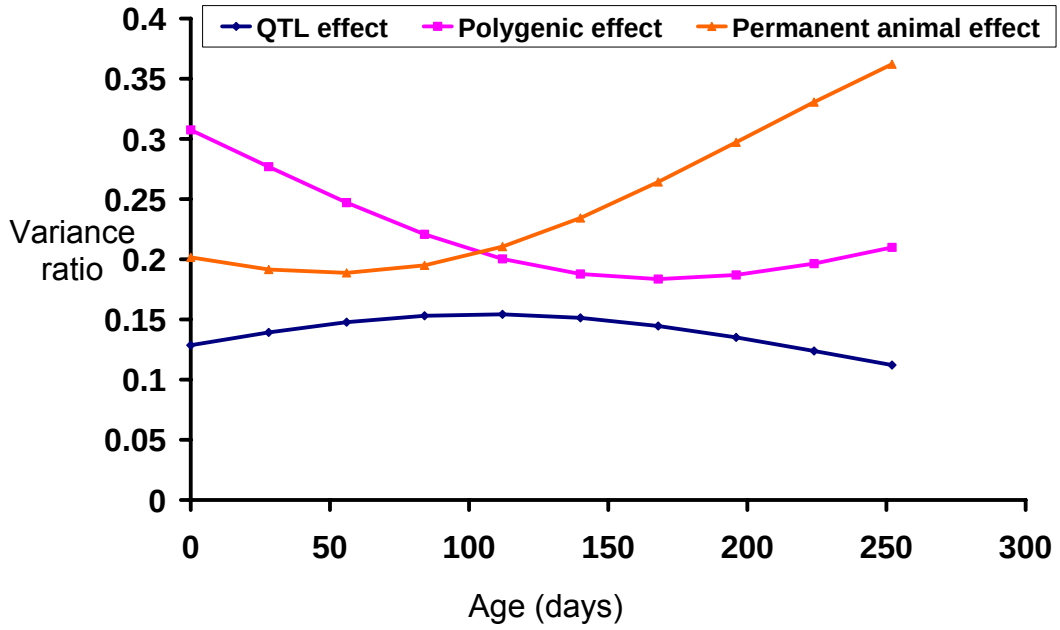
[1]NC: Model fit did not converge.

Appendix 6.5

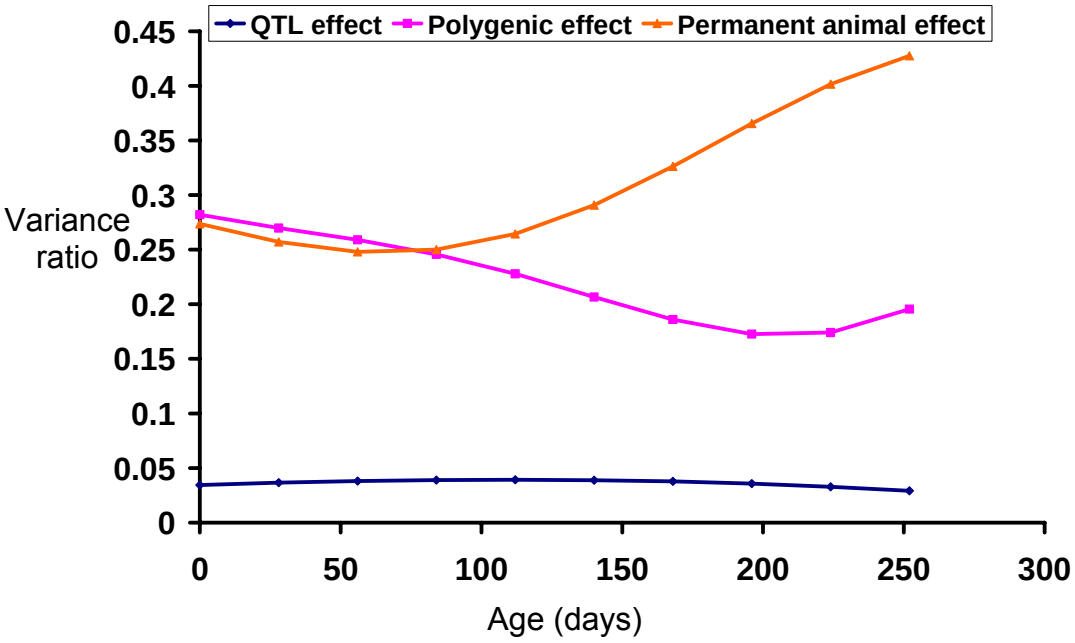
Variance ratios over time for all components fitted in statistically suboptimal models for selected scenarios.

Live Weight analysis

$QTL_{constant}$ scenario: Full model with zero order for QTL and 1st order for other effects. Variance ratios predicted at 8 cM.

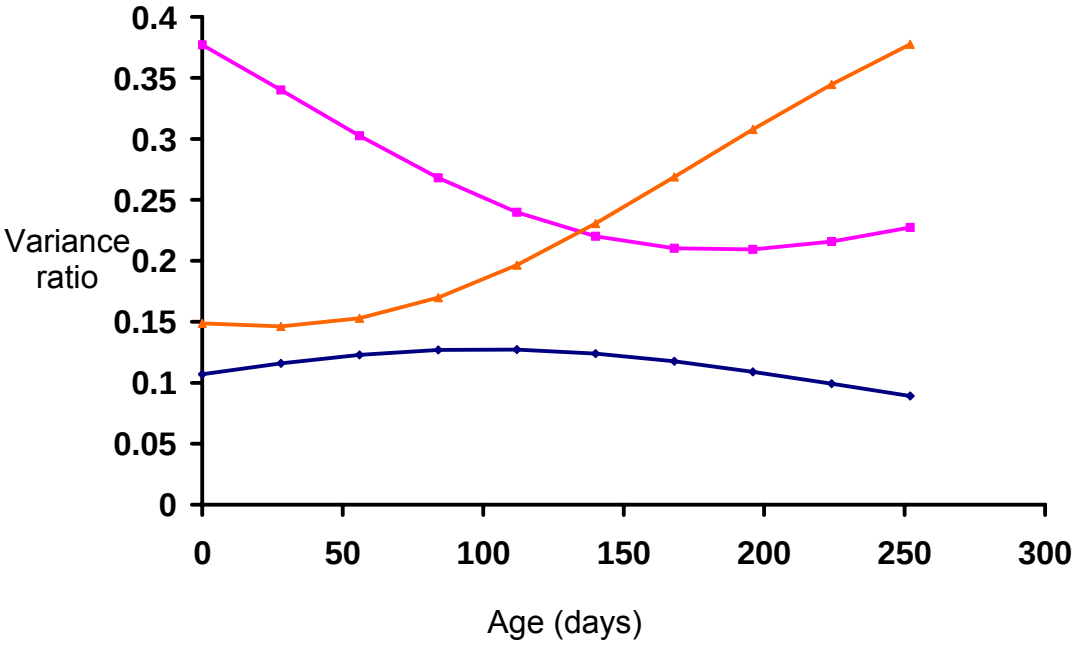


$QTL_{midpoint}$ scenario: Full model with zero order for QTL, 1st order for permanent animal, and 2nd order for polygenic effect. Variance ratios predicted at 8 cM.

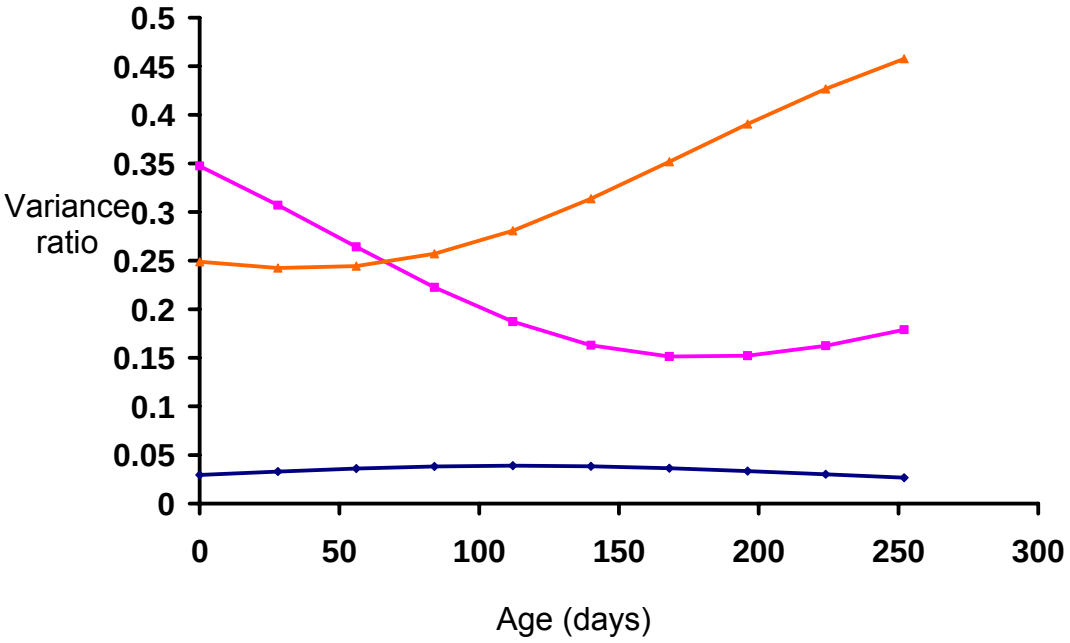


Daily gain analysis

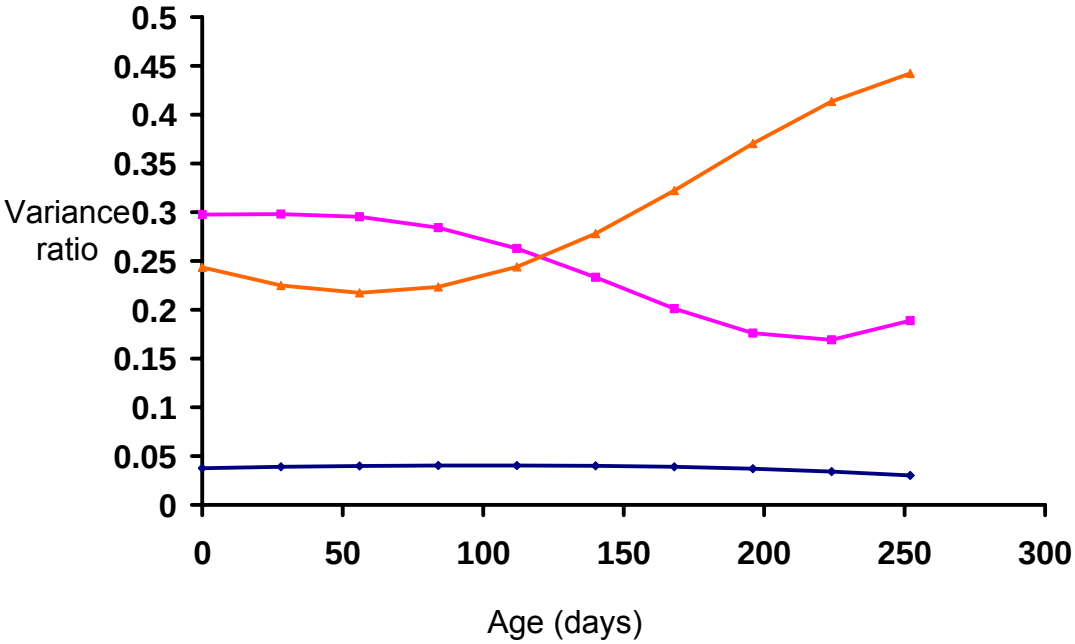
$QTL_{constant}$ scenario: Full model with zero order for QTL and 1st order for other effects. Variance ratios predicted at 9 cM.



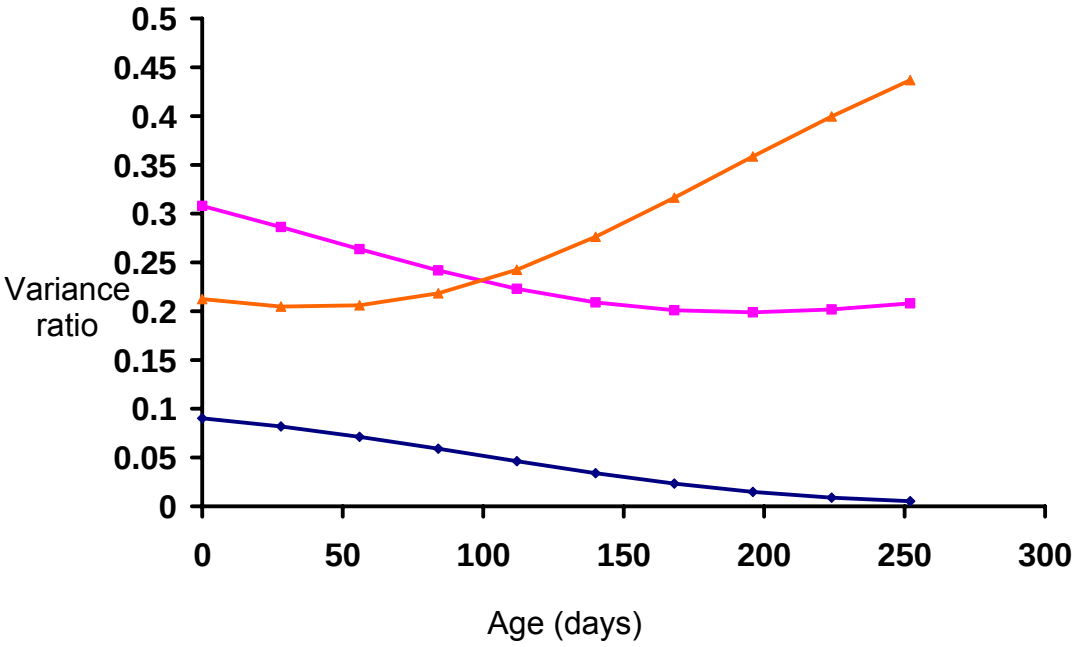
QTL_{high_low} scenario: Full model with zero order for QTL and 1st order for other effects. Variance ratios predicted at 7 cM.



$QTL_{midpoint}$ scenario: Full model with zero order for QTL, 1st order for permanent animal and 2nd order for polygenic effect. Variance ratios predicted at 9 cM.



$QTL_{midpoint}$ scenario: Full model with 1st order for QTL, permanent animal and polygenic effects. Variance ratios predicted at 9 cM.



CHAPTER 7

General Discussion

7.1 Thesis motivation

Growth and muscle composition are key traits that shape lamb performance and, thus, are of primary interest to sheep breeders. Selection of animals for breeding based on genetic loci that directly influence or are linked to causative variant(s) affecting such economically important traits presents a valuable strategy for the sheep industry. With the work described in this thesis, I endeavoured to detect and extensively characterise genetic aspects affecting muscle and growth phenotypes in sheep. In conjunction with the above, I sought to explore and develop approaches for efficient description of genetic components of traits that vary with time, such as growth. The only longitudinal trait studied in this thesis was growth, with primary focus being on describing genetic aspects of sheep growth. Nonetheless, the general incentive was to examine methods that would be applicable across livestock species and on approaches whose underlying principles would be pertinent to other longitudinal traits.

7.2 Thesis objectives

Firstly, the study aimed at determining whether the same genetic locus was responsible for previously mapped QTL for weight, muscle and fat traits on ovine chromosome 2 (OAR2) in British commercial breeds (Walling *et al.* 2004; McRae *et al.* 2005). For this purpose, a genetic study of two *GDF8* single nucleotide polymorphisms (SNPs), identified in progeny of Belgian Texel rams exhibiting muscular hypertrophy (Clop *et al.* 2006), was pursued in commercial Suffolk, Texel, and Charollais populations. After determining whether the polymorphisms were present in one or more of the breeds, the principal objective was to conduct association analysis of the SNP effects on muscle and fat phenotypes. This analysis would further evaluate and characterise the contributions of SNP effects to trait differences, with the findings being directly relevant to the sheep industry.

Secondly, the research focused on dissecting genetic aspects of growth as a process in sheep. The specific aim was to identify and describe QTL that capture genetic variability in longitudinal live weights at diverse points in time. Hence, a novel method was developed for longitudinal trait analysis and assessed for growth QTL detection. Additionally, applicability of a second, previously proposed method was explored. The first method employed descriptors of growth derived from fitted growth curves for QTL mapping and characterisation. The second approach aimed at assessing the application of random regression (RR) models for dissection of time-dependent QTL. Analyses of both field and simulated phenotypes were pursued with the aforementioned methods.

7.3 Overview of outcomes

As it was described in **chapter 2**, the favourable *g-2449C* and *g+6723A* alleles at the two loci studied were absent from commercial Suffolk and nearly fixed in the Texel breed, or at least in the sample examined. Each of the favourable alleles, however, was found to segregate with a frequency of ~0.3 in the Charollais breed. Mixed model SNP association analyses using extensive phenotypic and pedigree records for Charollais sheep revealed that both loci had a significant association with ultrasound scanned muscle depth but not with fat depth or weight at scanning. Additionally, it was shown that the two loci exhibited substantial linkage disequilibrium (LD). The findings supported the notion that the real functional effect on muscle depth arose from the *g+6723A* allele, as proposed by Clop *et al.* (2006), and that the detected effect of the *g-2449C* allele was due to its LD with the causative variant.

With respect to the *g+6723G>A* locus, animals with the AA genotype had significantly greater muscle depth than those with either the GG or AG genotype. The A allele, the likely causative mutation, was estimated to have an additive effect of 1.20 (±0.30) mm and a dominance effect of -0.73 (±0.36) mm, indicating a partial recessive action on muscle depth. Based on estimated allelic effects and allele frequencies, the *g+6723G>A* locus explained 14% of the additive genetic variance of muscle depth in the Charollais breed.

Chapter 3 described a novel method for dissecting age-dependent growth QTL and examining changes in QTL effects over time. This method entailed choosing a growth model that provided optimal fit for live weight measurements

across time, estimating the model parameters and predicting growth traits at different time points for each individual. QTL were mapped for estimated model parameters and weekly-predicted live weights and growth rates. QTL on different chromosomes were found to be significant at distinct growth stages. Additionally, QTL significance and effects were shown to vary with age, with QTL for growth rate often occurring earlier than the equivalent QTL for live weight.

The growth curve approach for QTL mapping of longitudinal growth measurements was further validated in **chapter 4** using an independently simulated dataset of yield. QTL mapping using half-sib analyses of individual Gompertz curve parameter estimates, predicted yields and growth rates succeeded in capturing 16 out of 18 simulated QTL with one false positive. Overall, half-sib regression analysis of model-predicted yields and growth rates performed better in identifying the simulated QTL effects than linkage analysis of univariate phenotypes.

Furthermore, the growth model method was extended to perform genomic evaluation of longitudinal yield. Prediction of phenotypic information from the Gompertz curve using either of two methods detailed in **chapter 4**, made it possible to obtain genomic breeding value estimates for a time point with no phenotypic records (time 600). This was achieved for individuals with trait information at other time points (training set) and individuals with no phenotypes (candidate set). Retrospective comparison revealed that the accuracies between true and estimated breeding values at time 600 were above 0.93 for both training and candidate sets.

In **chapter 5**, an alternative approach for identifying QTL for longitudinal traits was explored using the same live weight data studied in **chapter 3**. Specifically, a systematic procedure was investigated for selecting RR models that

included a QTL as time-dependent random effect, in addition to time-varying polygenic and environmental effects. RR analysis of untransformed live weights resulted in the detection of apparently significant QTL on all four chromosomes studied (OAR5, 14, 18 and 20) when a RR polynomial of 2nd order or higher was fitted for the QTL effect. With the chosen full models, similar QTL variance trajectories over time were predicted for all chromosomes with significant QTL. These findings did not agree with the distinct patterns of expression for age-dependent QTL identified on the corresponding chromosomes using the growth model approach (**chapter 3**).

In **chapter 6**, dissection of longitudinal trait components with RR was further examined. For this purpose, four biologically relevant scenarios of growth were simulated and analysed using RR models. Irrespective of QTL time-dependence scenario, full models succeeded in identifying the simulated QTL as significant when this effect was fitted with low RR polynomial orders in either daily gain or live weight analyses. In all cases, apart from analyses of a QTL with maximum effect mid growth, models with a RR order for the QTL that effectively described the pattern of QTL variance over time were deemed statistically better than reduced models. Additionally, model-predicted trajectories along time generally agreed with expected profiles for the simulated polygenic and permanent animal effects. Overall, analyses of simulated data provided more comprehensible results than the study of Blackface live weights presented in **chapter 5**, with respect to using RR for dissecting QTL significance and predicting changes of variance components over time.

7.4 Conclusions and relevance of findings

Chapter 2 presented, to my knowledge, the first report of significant association of two *GDF8* polymorphisms with muscle depth in commercial Charollais sheep. The Charollais appears to be the only breed, apart from the Australian White Suffolk (Kijas *et al.*, 2007), in which the favourable *g+6723A* allele has an intermediate frequency. Additionally, in accordance with the strategy outlined by Ron & Weller (2007), the analyses of the locus effects provided further statistical validation for the *g+6723G>A* SNP being a quantitative trait nucleotide (QTN) for muscularity in sheep, as previously proposed by Clop *et al.* (2006). Further, a partially recessive effect of the QTN on muscle depth was suggested for the first time in **chapter 2**. Finally, a proposition was made regarding how the proposed molecular mechanism by which the *g+6723A* allele leads to increased muscle (Clop *et al.* 2006) accommodates partially recessive action of the mutation on myostatin expression. These findings may be used directly in breeding programmes.

The concept of fitting parametric curves to longitudinal data to estimate biologically relevant curve parameters had been previously put forward by various researchers (e.g. Lopez *et al.* 2000; Lewis *et al.* 2002; Schinckel *et al.* 2004; Lambe *et al.* 2006; Rodriguez-Zas *et al.* 2002; Minvielle *et al.* 2006). For example, Lewis *et al.* (2002) and Lambe *et al.* (2006) investigated the polygenic components of growth curve parameters whereas Rodriguez-Zas *et al.* (2002) and Minvielle *et al.* (2006) used parameter estimates to detect QTL for milk traits in dairy cattle and egg production in quail, respectively. The growth model procedure described in **chapter 3** and further validated in **chapter 4** represents an advancement of previous

applications. Work in both chapters provided a detailed description of growth model fitting for QTL mapping. Weekly growth rate and weight predictions were employed for the first time to map QTL, and, more importantly, to examine trends in QTL significance and variance over time. The outcomes underscored the value of the approach as a multivariate method for extracting more phenotypic information from longitudinal growth data, and decomposing multiple measurements into more informative, biologically meaningful variables for QTL analyses. Specifically: a) growth model descriptors resulted in increased ability to detect growth QTL as compared to univariate analyses of actual weight data; b) QTL effects were shown to be time-dependent; c) different patterns were found for QTL on different chromosomes, with individual growth QTL describing distinct parts of an animal's growth curve trajectory.

In addition, **chapter 4** also introduced a robust approach for genomic evaluation for a longitudinal trait. Performing separately the growth modelling and genomic evaluation processes led to great simplification of the evaluation methodology with no detrimental effects on final results. Furthermore, the method of using genomic estimated breeding values (GEBVs) of the three model parameters to evaluate the Gompertz function in order to estimate GEBVs for the trait at a time of interest could be beneficial when GEBVs are needed for growth at different time points, by simplifying the estimation process. Moreover, the fact that the proportion of SNPs affecting the trait was estimated from the data for the first time, in the aforementioned study, constitutes further development of the BayesB approach for genomic evaluation.

RR models, although applied in many occasions to model polygenic, maternal genetic and environmental components for various longitudinal traits (e.g. Meyer 1998; Lewis & Brotherstone 2002; Fischer *et al.* 2006; Kranis *et al.* 2007), have only recently been proposed for identification of time-dependent QTL (Lund *et al.* 2002; Macgregor *et al.* 2005; Lund *et al.* 2008). Analyses of both simulated Macgregor *et al.* 2005; Lund *et al.* 2008) and actual phenotypes (Lund *et al.* 2008) have indicated that even low RR orders can lead to increased power for QTL detection, when the QTL has a sizeable effect at a specific time point (or range) while its variance changes substantially over time. In the above studies, fitting higher order RR polynomials for the QTL resulted in model convergence problems. Findings from analyses of simulated live weights in **chapter 6** tended to agree with both of the aforementioned results.

When actual live weights were analysed using RR in **chapter 5**, model convergence at local maxima seemed to occur even prior to fitting a QTL effect. QTL mapping outcomes from selected models were not in accordance with results from the growth model approach described in **chapter 3** and findings from QTL mapping using moving average phenotypes (results not shown). Thus, the RR results were inconclusive regarding statistical significance and time-dependence of QTL on different chromosomes. Overall, with the given data structure, the RR approach had a poorer ability to describe growth QTL over time than the growth model procedure.

As a whole, the results from both **chapters 5 and 6** emphasised the importance and complexity of choosing the optimal RR model to dissect time-dependent QTL. Judicious modelling has long been deemed as crucial in the RR framework (Meyer 1998). In the current study, it was shown that for QTL mapping

in particular, it is critical to account for all other relevant time-dependent variance components in full models. Only then would RR models allow both the detection of significant QTL and ostensibly reliable predictions of the QTL variance trajectory over time. However, even in the limited number of simulation scenarios explored in **chapter 6**, QTL variance predictions followed the expected pattern only when the change in QTL expression was straightforward, i.e. gradual and roughly linear over time, and when no confounding with the other time-dependent effects was apparent. Unless these conditions can be assumed *a priori* in practice, caution is advised for relying on the characterisation of age-dependent QTL solely from statistically optimal RR models. After exhaustive checks to find the most relevant yet parsimonious model, it would be necessary to specify certain criteria for assessing or determining that the statistically sound model would also be biologically optimal prior to committing to the QTL results.

Finally, assessing the performance of the two methods for dissecting QTL for longitudinal traits led to the conclusion that the approximate parametric growth model approach seems to be a simple and robust method, whereas the more statistically rigorous procedure of random regression appears to be lacking for mapping time-dependent QTL, at least under the data structures investigated.

7.5 Challenges and perspectives for future research

Work presented in this thesis provided substantial insight into the suitability of the growth model and random regression approaches for mapping time-dependent QTL for a longitudinal trait. In addition, the thesis revealed challenges posed mainly by the attributes of each approach, as well as opportunities for future research.

With respect to the growth model method, issues about its application originate directly from growth curve properties. The first concern arises from the fact that, as was mentioned in **chapter 5**, growth curves combine live weight data only at the phenotypic level and they constrain variation around a set phenotypic structure over time. Thus, information arising from patterns of change in the trait's genetic and environmental components that do not conform to the curve trajectory may be missed (Meyer & Kirkpatrick 2005). This issue has not been addressed in the thesis and could be studied further, possibly using simulations. Comparative assessment of the growth model and RR methods, with the latter one providing more flexible fit to the data, indicated that missing variation not accounted for by the fitted growth curve may be a small price to pay. This is because overall robustness of the growth curve approach for mapping QTL was perceived, with choice and application of the optimal RR model for the same task being particularly complex.

Another potential concern that has been previously raised about growth curves is that estimates of curve parameters may be highly correlated (Cullis & McGilchrist 1990; Lewis *et al.* 2002), although the true parameter values might not be (Lewis & Brotherstone 2002). In **chapter 3**, different QTL were mapped for the three parameters estimated from the Gompertz curve, indicating that high phenotypic correlations between their estimates may not necessarily translate into high genetic

correlations. However, dissection of the majority of time-dependent QTL was possible when curve predictions of growth traits, rather than curve parameters, were analysed, indicating that the former descriptors are more informative for QTL detection.

The work described in **chapter 4** provided information regarding both above-mentioned issues. The fact that fitting the Gompertz model to data simulated using the logistic function led to successful QTL mapping and genomic evaluation of the trait at a time point beyond the data range suggested that: a) choosing the model that provides good fit for the data is of greater importance than the constraints imposed to phenotypic variation by the parametric curve fitted, and b) using trait predictions from the growth curve to identify QTL may be a better strategy for determining genetic merit for animals than using the estimated curve parameters.

Another possible challenge for growth curve modelling and not for RR, which was also stated in previous studies, rests on the fact that differences exist in the precision of estimating model parameters between animals, particularly when different numbers of live weight records are available per individual, as is often the case in practice (Lewis *et al.* 2002; Lewis & Brotherstone 2002). This limitation has not been addressed thus far. It could be argued that it is minimised by using similar number of records, measured close in time for all animals, as was done in this thesis. This would lead to greater probability of an approximately uniform measurement error across animals at each time point. However, further research would be required to assess the consequences of estimation errors on the precision of QTL mapping using model parameter estimates and trait predictions.

In addition to parameter estimation errors, temporary environmental fluctuations, such as gut fill, can have a major impact on the weight value recorded per animal at different time points or even intervals in a given day. Thus, increased measurement "noise" would be expected to have considerable effects on univariate analyses of weight for QTL detection. Although temporary environmental effects are not explicitly accounted for by growth curves, the "noise" is likely averaged out. This expectation arises from the fact that an animal's growth trajectory gets smoothed by fitting a growth curve to weight records. Thus, curve predictions would be less influenced by temporary fluctuations than actual records. Therefore, in cases that substantial temporary fluctuations influence the recorded phenotypes, employing growth model predictors would be more advantageous for identification of growth QTL than univariate analyses of actual weights.

The novel approach presented in **chapter 4** for genomic evaluation of a longitudinal trait presents great opportunities for future research. Mainly, it could be employed and developed within the realm of the latest technologies combining quantitative genetic principles with advances in molecular biology. Specifically, the proposed method may facilitate efficient application of a) whole genome evaluation and b) genome-wide association studies (GWAS) on longitudinal traits.

Extension of current genomic evaluation methods to multivariate analyses of SNP effects has not been reported thus far. Such development may not be straightforward as the method should accommodate the fact that individual SNPs may be affecting all, some or none of the traits included in the analysis. Additionally, current methods of evaluation do not account for time-varying SNP effects to provide effective description of longitudinal traits. Yet, evaluation of animals at

various stages of growth, lactation, egg production, or for other longitudinal traits, would be necessary since such traits seem to be governed by distinct genetic components at different times. Further, evaluation of single time point records for each trait may be misleading as it may not allow dissection of significant effects for some loci. As was shown in **chapter 4**, genomic evaluation for time-varying traits can potentially be achieved using a simpler approach than a full multivariate analysis. It could be pursued by fitting the optimal parametric curve for a given longitudinal phenotype, and using biologically relevant parameter estimates and informative trait predictions at various time points as traits. Another implication of the results of this method is that genomic evaluation may be successfully pursued at a time point for which no phenotypic information is available. This could potentially expedite the procedure; for example, by using trait predictions obtained from the relevant curve, young animals could be evaluated for a predicted trait at a time point in maturity, long before they reach that age.

In addition to genomic evaluation, parametric curve descriptors may be useful in GWAS. These could be analysed to identify single nucleotide (or other) polymorphisms that may display time-dependent changes in significance for a longitudinal trait. In GWAS, the variant would be fitted as a fixed effect in the models tested, as is currently done in studies of traits at a single point in time. The benefit would arise from analysing model-predicted trait values at distinct time points and assessing the respective effect of the polymorphism.

Also, depending on trait characteristics, genome-wide studies often require great amounts of data even for single time point analyses to allow for moderate detection power and to reduce false positive results. Power and false discovery

parameters may also be influenced by other factors, such as the effective population size, minor allele frequencies, study design and the significance thresholds adopted. Under the same conditions used for univariate analyses, it may be hypothesized that employing more informative descriptors of longitudinal traits would lead to increased power for detecting significant variants, particularly if significance is time-dependent.

It is important to note though that to obtain reliable curve predictions for a trait over time, multiple phenotypes over time need to be recorded for a given number of animals. Thus, data and population structure requirements for longitudinal trait GWAS or genomic evaluation need to be investigated further to balance practical difficulties with benefits. Based on this, decisions regarding the usefulness of the method may be trait-dependent. Nevertheless, it is important to account for the time component when attempting to identify genetic variants and QTL for longitudinal traits, especially in large scale, whole genome experiments.

The RR approach, studied in **chapters 5** and **6**, has previously been proposed as more powerful and more flexible than a growth model for analysing longitudinal live weights to determine genetic merit of animals (Lewis & Brotherstone 2002; Meyer & Kirkpatrick 2005). In part, this appears to be statistically valid, since RR solutions of genetic and environmental components are derived at the same time for all individuals. However, this advantage rests upon certain assumptions, such as the following: that the statistically best RR model can easily be chosen and that it will also be biologically appropriate. Many studies to date have shown that RR methods model the polygenic and environmental components of live weights over time more accurately than univariate analyses (e.g. Meyer *et al.* 1998; Fischer *et al.* 2004b;

Schaeffer 2004), even with statistically suboptimal models (Meyer 1998). Yet, specifically for describing growth QTL over time, the current study indicated that complications in the implementation of the method and in model choice challenge potential effectiveness of RR models.

Model choice was found to be particularly challenging for analysis of actual longitudinal live weights in this thesis. To choose the optimal model, the recommendations of Meyer (1998) were followed, specifically about allowing the data to determine the order of polynomial fit required. Yet, even in the models that were deemed statistically optimal with this procedure, several time-dependent variance components were included, potentially resulting in overparameterisation. Model overfitting may also have been a result of inefficient study design, in terms of population structure and/or amount of data. This has been proposed previously (Meyer *et al.* 1998; Molina *et al.* 2007). However, with respect to the sheep industry, the available dataset is representative of commercial population structures and the number of live weight records and genotyped animals that could conceivably be obtained by breeders, without genotyping and recording costs being unrealistic. Although the growth curve approach seemed to efficiently analyse this dataset, choice of the RR model that would best describe these records seemed problematic even prior to including a QTL effect in the models.

While implementation of RR may have been hindered by a potentially limiting dataset in **chapter 5**, this has not been investigated in the current study, which focused on assessing how effective RR is in describing QTL effects that vary with time. Although the power of the method was not examined in the simulation study presented in **chapter 6**, better population structure and data quality may have

contributed to increased effectiveness of RR models in modelling the time-varying effects for simulated growth traits. In future work, the power of RR for identifying QTL, in addition to modelling other time-dependent components, needs to be investigated for various amounts of phenotypes and various population structures in future studies. RR analyses of biologically simplified simulation scenarios, performed by Macgregor *et al.* (2005) and Lund *et al.* (2008), have indicated increased power for QTL detection compared to univariate analyses and repeatability models fitted to the same datasets.

Data structure has likely not been the only source of complications in model choice. Restrictions on the polynomial orders fitted may have also been imposed by the intrinsic properties of the RR method. Convergence problems along with erratic variance predictions, particularly at the edges of the age range, were apparent in the study of Blackface live weights when higher orders of Legendre polynomials were fitted. These outcomes agreed with previous studies, which had found that higher order polynomials magnified small sampling errors (Meyer 1998; Meyer 1999; Fischer & van der Werf 2002; Fischer *et al.* 2004a, b) and that RR models displayed overall sensitivity with respect to effects and orders of covariance functions fitted (Meyer 1999; Meyer & Kirkpatrick 2005). Additionally, although the use of Legendre polynomials for RR has been favoured for their comparative advantages over other polynomials, Legendre functions seem to place weight on records at the extremes of the time interval for which they are defined and in which the polynomials are most unstable (Meyer 2001; Fischer *et al.* 2004a).

Aside from the above limitations of polynomials, Macgregor *et al.* (2005) had suggested only testing low order polynomials, unless a great amount of data is

available, even in cases where higher orders result in improved fit. The argument of Macgregor *et al.* (2005) rested on the fact that higher orders of fit also require estimation of more parameters. This greatly complicates convergence and statistical testing and, thus, assessment of QTL significance. Although this approach seems reasonable, it must be noted that the QTL effect correlation structure and change in variance over time are specified together with the RR approach. Thus, although a low order polynomial may adequately model variance change over time, it may not accommodate the covariance structure or vice versa (Macgregor *et al.* 2005). In this thesis, QTL variance prediction over time using RR was indeed found to be more challenging than assessment of QTL significance, in particular when using field data (**chapter 5**). Thus, the implications of fitting suboptimal polynomial orders may be more subtle for describing QTL effects than other time-dependent effects. This disadvantage of the RR method, that may prevent accurate QTL variance prediction over time, remains an important issue that should be examined in the future. Alternatively, for the description of time-dependent QTL, it may be necessary to employ methods that fit distinct functions to model changes in variance and covariance (or correlation) over time.

Computational time has also been a major issue when testing RR methodology in the current study and in literature (e.g. Meyer 1998; Meyer & Kirkpatrick 2005). Testing multiple polynomial fits while scanning along each chromosome in **chapter 5** was extremely time consuming, even when using the powerful computing facilities of the University of Edinburgh (ECDF). One of the aims of the two-step approach tested for RR model choice was to minimise computational time. This would presumably be achieved by conducting a small

number of genome scans by fitting varying orders for the QTL effect to pre-selected optimal models for all other effects. Although this approach was helpful, computational time remained a rate-limiting factor. The recommendations given by Macgregor *et al.* (2005) could serve as additional guidance in order to address this. It is important though to keep in mind that, as was shown in this thesis, accounting for all relevant time-dependent components in the RR model seems necessary. This is to ensure efficient variance partitioning and description over time when attempting to map QTL, even if it further complicates model choice.

The complexity of determining the likely mixture distributions of the likelihood ratio test statistics under the null hypothesis to assess the significance of fitting additional time-dependent effects in RR models has been described for some cases by Meyer (1998). This complication was recently reiterated for determining time-dependent QTL significance by Lund *et al.* (2008). Having this in mind, the use of a naïve chi-squared distribution, as was done by Lund *et al.* (2008) and in the current study, is likely too conservative. However, this choice was based on the fact that the mixture of chi-square distributions is not known when fitting time-dependent QTL with RR models and that any less conservative test statistic may have led to higher chances of obtaining falsely significant QTL. Further research is required to determine the appropriate test statistics for including a QTL effect and increasing the RR orders to model longitudinal QTL. Using permutation testing, although recommended for setting significance thresholds in QTL mapping methodologies, would be exceptionally demanding computationally in the RR framework (Lund *et al.* 2008).

Future research on methods other than the growth model approach for time-dependent QTL identification should focus on more parsimonious models. Most approaches have been previously suggested or tested for analysis of time-varying polygenic and environmental components. Procedures within the RR framework could correspond to: a) reduced rank covariance matrices for each time-dependent effect (Meyer 1998; Meyer 1999; Meyer & Kirkpatrick 2005); b) principal component analysis of the covariance functions (Kirkpatrick & Meyer 2004; Meyer & Kirkpatrick 2005) and c) fitting splines instead of continuous polynomial functions (e.g. Verbyla *et al.* 1999; White *et al.* 1999; applications summarised in Meyer & Kirkpatrick 2005). Alternatively, character process models (Pletcher & Geyer 1999; Jaffrezic & Pletcher 2000) and structured antedependence models (Meyer 2001; Jaffrezic et al. 2003) could be examined. Overall, findings of this thesis regarding model choice using any methodology to map time-dependent QTL, lead to recommendations that are in line with those made by Thompson *et al.* 2005. In particular, along with appreciation of the method's properties and power, it is necessary for alternative models to be compared using genetic and biological measures rather in addition to statistical selection criteria.

7.6 Implications and practical considerations

Two broad practical implications of the thesis findings can be proposed. Firstly, the significant association between the two *GDF8* SNPs with increased muscle depth indicates that marker assisted selection using these SNPs would be beneficial for an additional sheep breed; the Charollais. Specifically, the estimated maximum additive genetic variance for muscle depth that could be attributed to the *g+6723G>A* locus would be attained at a *g+6723A* allele frequency of 0.7 and would equal 38% of the total additive genetic variance of muscle depth. Thus, selection for this locus seems highly advantageous for the breeders.

Moreover, given the indication that the favourable *g+6723A* allele has partially recessive action on muscle phenotype, the rate of genetic progress for the trait would depend heavily not just on the allele frequency but also on the proportion of homozygous animals for this variant in the population. Consequently, methodically planned marker assisted selection (MAS) for this locus would be better than mass selection for ensuring that a significant increase in frequency of the favourable genotype would occur in a given time. This is particularly relevant when, for instance, sire referencing schemes are used for breeding, as is done in Great Britain.

Since the *g+6723A* allele has been shown to be initiating a molecular pathway for myostatin inhibition, it is directly connected with muscle production. Thus far, no allelic effects on meat quality traits have been reported for New Zealand Texel sheep (Johnson *et al.* 2009), or Belgian Texel progeny in which this polymorphism was first characterised (Clop *et al.* 2006). Nevertheless, given potential background effects in different breeds, the association of the QTN with

meat quality measurements could be investigated further in the Charollais breed to ensure informed MAS with no adverse side effects for the breeders.

A second practical implication from thesis outcomes rests on the detection of QTL that describe distinct parts of an animal's growth curve trajectory. This presents potential opportunities for manipulation of the growth trajectory based on the breeders' selection objectives. This is relevant not only for sheep but also for other livestock industries. Additionally, further development of the proposed method for identifying and describing time-dependent polymorphisms with significant effects on growth (as was explained above) may, in the future, allow comprehensive and genome-wide screening for marker assisted selection and more accurate genomic evaluation for different aspects of longitudinal growth traits.

BIBLIOGRAPHY

Alfonso L. & Haley C.S. 1998. Power of different F_2 schemes for QTL detection in livestock. *Animal Science* **66**, 1-8.

Allison D.B., Neale M.C., Zannolli R., Schork N.J., Amos C.I. & Blangero J. 1999. Testing the robustness of the likelihood-ratio test in a variance-component quantitative-trait loci-mapping procedure. *American Journal of Human Genetics* **65**, 531-544.

Ap Dewi I., Saatci M. & Ulutas Z. 2002. Genetic parameters of weights, ultrasonic muscle and fat depths, maternal effects and reproductive traits in Welsh Mountain sheep. *Animal Science* **74**, 399-408.

Apiolaza L.A., Gilmour A.R & Garrick D.J. 2000. Variance modelling of longitudinal height data from a *Pinus radiata* progeny test. *Canadian Journal of Forest Research* **30,** 645-654.

Australian sheep gene mapping web site: http://rubens.its.unimelb.edu.au/~jillm/jill.htm

Barrett J.C., Fry B., Maller J. & Daly M.J. 2005. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.

Bignardi A.B., El Faro L., Cardoso V.L., Machado P.F. & Albuquerque L.G. 2009a. Parametric correlation functions to model the structure of permanent environmental (co)variances in milk yield random regression models. *Journal of Dairy Science* **92**, 4634-4640.

Bignardi A.B., El Faro L., Cardoso V.L., Machado P.F. & Albuquerque L.G. 2009b. Random regression models to estimate test-day milk yield genetic parameters Holstein cows in Southeastern Brazil. *Livestock Science* **123**, 1-7.

Boman I.A., Klemetsdal G., Blichfeldt T., Nafstad O. & Våge D.I. 2009. A frameshift mutation in the coding region of the *myostatin* gene (*MSTN*) affects carcass conformation and fatness in Norwegian White Sheep (*Ovis aries*). *Animal Genetics* **40**, 418-422.

Boman I.A. & Våge D.I. 2009. An insertion in the coding region of the *myostatin* (*MSTN*) gene affects carcass conformation and fatness in the Norwegian Spælsau (*Ovis aries*). *BMC Research Notes* **2**, 98.

Broad T.E., Glass B.C., Greer G.J., Robertson T.M., Bain W.E., Lord E.A. & McEwan J.C. 2000. Search for a locus near to *myostatin* that increases muscling in Texel sheep in New Zealand. *Proceedings of New Zealand Society of Animal Production* **60**, 110-112.

Brotherstone S., White I.M.S. & Meyer K. 2000. Genetic modelling of daily milk yield using orthogonal polynomials and parametric curves. *Animal Science* **70**, 407-415.

Bünger L. & Herrendörfer G. 1994. Analysis of a long-term selection experiment with an exponential model. *Journal of Animal Breeding and Genetics-Zeitschrift für Tierzüchtung und Züchtungsbiologie* **111**, 1–13.

Bünger L., Ott G., Varga L., Schlote W., Rehfeldt C., Renne U., Williams J.L. & Hill W.G. 2004. Marker assisted introgression of the Compact mutant myostatin allele: *Mstn(^Cmpt-dl1Abc)* into a mouse line with extreme growth effects on body composition and muscularity. *Genetical Research* **84**, 161-173.

Churchill G.A. & Doerge R.W. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963–971.

Clop A., Marcq F., Takeda H., Pirottin D., Tordoir X., Bibe B., Bouix J., Caiment F., Elsen J.M. Eychenne F., Larzul C., Laville E., Meish F., Milenkovic D., Tobin J., Charlier C. & Georges M. 2006. A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. *Nature Genetics* **38**, 813-818.

Cockett N.E., Jackson S.P., Shay T.L., Nielsen D., Moore S.S., Steele M.R., Barendse W., Green R.D. & Georges M. 1994. Chromosomal localization of the *callipyge* gene in sheep (*Ovis aries*) using bovine markers. *Proceedings of the National Academy of Science* **91**, 3019-3023.

Cockett N.E., Smit M.A., Bidwell C.A., Segers K., Hadfield T.L., Snowder G.D., Georges M. & Charlier C. 2005. The callipyge mutation and other genes that affect muscle hypertrophy in sheep. *Genetics Selection Evolution* **37**, S65-S81.

Corva P.M. & Medrano J.F. 2001. Quantitative trait loci (QTLs) mapping for growth traits in the mouse: A review. *Genetics Selection Evolution* **33**, 105-132.

Coster A., Bastiaansen J., Calus M., Maliepaard C. & Bink M. 2010. QTLMAS 2009: simulated dataset. *BMC Proceedings* **4**(Suppl. 1), S3.

Cullis B.R. & McGilchrist C.A. 1990. A model of the analysis of growth data from designed experiments. *Biometrics* **46**, 131-142.

Davies G., Stear M.J., Benothman M., Abuagob O., Kerr A., Mitchell S. & Bishop S.C. 2006. Quantitative trait loci associated with parasitic infection in Scottish blackface sheep. *Heredity* **96**, 252–258.

Fischer T.M., Gilmour A.R. & van der Werf J.H.J. 2004a. Computing approximate standard errors for genetic parameters derived from random regression models fitted by average information REML. *Genetics Selection Evolution* **36**, 363-369.

Fischer T.M. & van der Werf J.H.J. 2002. Effect of data structure on the estimation of genetic parameters using random regression. In: *7th World Congress on Applied Livestock Production*, Communication No. 17-08 (CD-ROM).

Fischer T.M., van der Werf J.H.J., Banks R.G. & Ball A.J. 2004b. Description of lamb growth using random regression on field data. *Livestock Production Science* **89**, **1**75-185.

Fischer T.M., van der Werf J.H.J., Banks R.G., Ball A.J. & Gilmour A.R. 2006. Genetic analysis of weight, fat and muscle depth in growing lambs using random regression models. *Animal Science* **82**, 13-22.

Gilbert H. & Le Roy P. 2003. Comparison of three multitrait methods for QTL detection. *Genetics Selection Evolution* **35**, 281-304.

Gilmour A.R., Gogel B.J., Cullis B.R., Welham S.J. & Thompson R. 2002. ASReml User Guide Release 1.0 VSN International Ltd, Hemel Hempstead, HP1 1ES, UK.

Green P., Falls K. & Crooks S. 1990. Cri-map version 2.4. Washington Univ. School Med., St. Louis, MO.

Grisart B., Coppieters W., Farnir F., Karim L., Ford C., Berzi P., Cambisano N., Mni M., Reid S., Simon P., Spelman R., Georges M. & Snell R. 2002. Positional candidate cloning of a QTL in dairy cattle: Identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research* **12**, 222-231.

Hadjipavlou G. & Bishop S.C. 2009. Age-dependent quantitative trait loci affecting growth traits in Scottish Blackface sheep. *Animal Genetics* **40,** 165-175.

Hadjipavlou G., Hemani G., Leach R., Louro B., Nadaf J., Rowe S. & De Koning D.J. 2010. Extensive QTL and association analyses of the QTLMAS2009 data. *BMC Proceedings* **4**(Suppl. 1), S11.

Heath S.C. 1997. Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. *American Journal of Human Genet*ics **61**, 748-760.

Hill W.G. & Robertson A. 1968. Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics* **38**, 226-231.

Jaffrézic F. & Pletcher S.D. 2000. Statistical models for estimating the genetic basis of repeated measures and other function-valued traits. *Genetics* **156**, 913-922.

Jaffrézic F., Thompson R. & Hill W.G. 2003. Structured antedependence models for genetic analysis of multivariate repeated measures in quantitative traits. *Genetical Research* **82**, 55-65.

Jamrozik J. & Schaeffer L. 1997. Estimates of genetic parameters for a test day model with random regressions for production of first lactation Holsteins. *Journal of Dairy Science* **80**, 762-770.

Janss L.L.G., Thompson R. & Arendonk J.A.M. van 1995. Application of Gibbs sampling for inference in a mixed major gene-polygenic inheritance model in animal populations. *Theoretical and Applied Genetics* **91**, 1137-1147.

Jiang C. & Zeng. Z.B. 1995. Multiple Trait Analysis of Genetic Mapping for Quantitative Trait Loci. *Genetics* **140**:1111-1127.

Johnson P.L., Dodds K.G., Bain W.E., Greer G.J., McLean N.J., McLaren R.J., Galloway S.M, van Stijn T.C. & McEwan J.C. 2009. Investigations into the GDF8 *g+6723G-A* polymorphism in New Zealand Texel sheep. *Journal of Animal Science* **87**, 1856-1864.

Johnson P.L., McEwan J.C., Dodds K.G., Purchas R.W. & Blair H.T. 2005. A directed search in the region of GDF8 for quantitative trait loci affecting carcass traits in Texel sheep. *Journal of Animal Science* **83**, 1988-2000.

Jones H.E., Lewis R.M., Young M.J. & Simm G. 2004. Genetic parameters for carcass composition and muscularity in sheep measured by X-ray computer tomography, ultrasound and dissection. *Livestock Production Science* **90**, 167-179.

Kambadur R., Sharma M. Smith T.P.L. & Bass J.J. 1997. Mutations in *myostatin* (*GDF8*) in double-muscled Belgian Blue and Piedmontese cattle. *Genome Research* **7**, 910-915.

Karamichou E., Richardson R.I., Nute G.R., Mclean K.A. & Bishop S.C. 2006. A partial genome scan to map quantitative trait loci for carcass composition, as assessed by x-ray computer tomography, and meat quality traits in Scottish blackface sheep. *Animal Science* **82**, 301-309.

Kijas J.W., McCulloch R., Edwards J.E.H, Oddy V.H., Lee S.H. & van der Werf J. 2007. Evidence for multiple alleles effecting muscling and fatness at the ovine GDF8 locus. *BMC Genetics* **8**, 80.

Kirkpatrick M., Lofsvold D. & Bulmer M. 1990. Analysis of the inheritance, selection and evolution of growth trajectories. *Genetics* **124**, 979-993.

Kirkpatrick M. & Meyer K. 2004. Direct estimation of genetic principal components: simplified analysis of complex phenotypes. *Genetics* **168**, 2295-2306.

Knott S.A., Elsen J.M. & Haley C.S. 1996. Methods for multiple marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71–80.

Knott S.A. & Haley C.S. 2000. Multitrait Least Squares for Quantitative Trait Loci Detection. *Genetics* **156**:899-911.

Knott S.A. 2005. Regression-based quantitative trait loci mapping: robust, efficient and effective. *Philosophical Transactions of the Royal Society B* **360**:1435-1442.

Koning D.J. de, Pong-Wong R., Varona L., Evans G.J., Giuffra E., Sanchez A., Plastow G., Noguera J.L., Andersson L. & Haley G.S. 2003. Full pedigree quantitative trait locus analysis in commercial pigs using variance components. *Journal of Animal Sci*ence **81**: 2155-2163.

Kranis A., Su G., Sorensen D. & Woolliams J.A. 2007. The application of random regression models in the genetic analysis of monthly egg production in turkeys and a comparison with alternative longitudinal models. *Poultry Science* **86**, 470-475.

Lambe N.R., Navajas E.A., Simm G. & Bünger L. 2006. A genetic investigation of various growth models to describe growth of lambs of two contrasting breeds. *Journal of Animal Science* **84**, 2642-2654.

Lee S-J. 2007. Quadrupling muscle mass in mice by targeting TGF-β signaling pathways. *PLoS ONE* **2(8)**, e789.

Lewis R.M. & Brotherstone S. 2002. A genetic evaluation of growth in sheep using random regression techniques. *Animal Science* **74**, 63-70.

Lewis R.M., Emmans G.C., Dingwall W.S. & Simm G. 2002. A description of the growth of sheep and its genetic analysis. *Animal Science* **74**, 51-62.

Lopez S., France J., Gerrits W.J.J., Dhanoa M.S., Humphries D.J. & Dijkstra J. 2000. A generalized Michaelis-Menten equation for the analysis of growth. *Journal of Animal Science* **78**, 1816-1828.

Lund M.S., Sorensen P. & Madsen P 2002. Linkage analysis in longitudinal data using random regression. In: *7th World Congress on Applied Livestock Production*, Communication No. 21-28 (CD-ROM).

Lund M.S., Sorensen P., Madsen P. & Jaffrézic F. 2008. Detection and modelling of time-dependent QTL in animal populations. *Genetics Selection Evolution* **40**, 177-194.

Ma C.X., Casella G. & Wu R. 2002. Functional mapping of quantitative trait loci underlying the character process: a theoretical framework. *Genetics* **161**, 1751-1762.

Macgregor S., Knott S.A., White I. & Visscher P.M. 2005. Quantitative trait locus analysis of longitudinal quantitative trait data in complex pedigrees. *Genetics* **171**, 1365-1376.

Maddox J.F., Davies K.P., Crawford A.M. *et al.* 2001. An enhanced linkage map of the sheep genome comprising more than 1000 loci. *Genome Research* **11**, 1275-1289.

Marcq F., Larzul C., Marot V., Bouix J., Eychenne F., Laville E., Bibé B., Leroy P.L., Georges M. & Elsen J.M. 2002. Preliminary results of a whole-genome scan targeting QTL for carcass traits in a Texel x Romanov intercross. *Proceedings of the 7th World Congress on Applied Livestock Production*, Montpelier, France, 19-23 August 2002, Abstract 02-14.

McPherron A.C. & Lee S.J. 1997. Double muscling in cattle due to mutations in the myostatin gene. *Proceedings of the National Academy of Science* **94**, 12457-1246.

McPherron A.C. & Lee S.J. 2002. Suppression of body fat accumulation in *myostatin*-deficient mice. *Journal of Clinical Investigation* **109**:595-601.

McRae A.F., Bishop S.C., Walling G.A., Wilson A.D. & Visscher P.M. 2005. Mapping of multiple quantitative trait loci for growth and carcass traits in a complex commercial sheep pedigree. *Animal Science* **80**, 135-141.

Meuwissen T.H.E., Hayes B.J. & Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-1829.

Meyer K. 1998. Estimating covariance functions for longitudinal data using a random regression model. *Genetics Selection Evolution* **30**, 221-240.

Meyer K. 1999. Estimates of genetic and phenotypic covariance functions for postweaning growth and mature weight of beef cows. *Journal of Animal Breeding and Genetics* **116**, 181-205.

Meyer K. 2001. Estimating genetic covariance functions assuming a parametric correlation structure for environmental effects. *Genetics Selection Evolution* **33**, 557-585.

Meyer K. 2002. Estimates of covariance functions for growth of Australian beef cattle form a large set of field data. In: *7th World Congress on Applied Livestock Production*, Communication No. 11-01 (CD-ROM).

Meyer K. & Kirkpatrick M. 2005. Up hill, down dale: quantitative genetics of curvaceous traits. *Philosophical Transactions of the Royal Society B* **360**, 1443-1455.

Minvielle F., Kayang B.B., Inoue-Murayama M., Miwa M., Vignal A., Gourichon D., Neau A., Monvoisin J.-L. & Ito S. 2006. Search for QTL affecting the shape of the egg laying curve of the Japanese quail. *BMC Genetics* **7**, 26.

Molina A., Menéndez-Buxadera A., Valera M. & Serradilla J.M. 2007. Random regression model of growth during the first three months of age in Spanish Merino sheep. *Journal of Animal Science* **85**, 2830-2839.

Nagamine Y., Haley C.S., Sewalem A. & Visscher P.M. 2003. Quantitative trait loci variation for growth and obesity between and within lines of pigs (*Sus scrofa*). *Genetics* **164**, 629-635.

Nicoll G.B., Burkin H.R., Broad T.E., Jopson N.B., Greer G.J., Bain W.E., Wright C.S., Dodds K.G., Fennessy P.F. & McEwan J.C. 1998. Genetic linkage of microsatellite markers to the Carwell locus for rib-eye muscling in sheep. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production* **26**, 529-532.

Pittroff W., Dahm F., Blanc F., Keisler D. & Cartwright T.C. 2008. Onset of puberty and the inflection point of the growth curve in sheep-Brody's Law revisited. *Journal of Agricultural Science* **146**, 239-250.

Pletcher S.D. & Geyer C.J. 1999. The genetic analysis of age-dependent traits: modelling the character process. *Genetics* **153**, 825-835.

Pong-Wong R., George A.W., Woolliams J.A. & Haley C.S. 2001. A simple and rapid method for calculating identity-by-descent matrices using multiple markers. *Genetics Selection Evolution* **33**, 453-471.

Pong-Wong R. & Hadjipavlou G. 2010. A two-step approach combining the Gompertz growth model with genomic selection for longitudinal data. *BMC Proceedings* **4**(Suppl. 1), S4.

R Development Core Team 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0; URL: http://www.R-project.org.

Rehfeldt C., Ott G., Gerrard D.E., Varga L., Schlote W., Williams J.L. & Bünger L. 2005. Effects of the myostatin mutation Compact on muscle cellularity in high growth mice. *Journal of Muscle Research and Cell Motility* **26**, 103-112.

Renne U., Langhammer M., Wytrwat E., Dietl G. & Bünger L. 2003. Genetic-statistical analysis of growth in selected and unselected mouse lines. *Journal of Experimental Animal Science* **42**, 218–232.

Riggio V., Finocchiaro R. & Bishop S.C. 2008. Genetic parameters for early lamb survival and growth in Scottish Blackface sheep. *Journal of Animal Science* **86**, 1758-1764.

Rodriguez-Zas S.L., Southey B.R., Heyen D.W. & Lewin H.A. 2002. Detection of quantitative trait loci influencing dairy traits using a model for longitudinal data. *Journal of Dairy Science* **85**, 2681-2691.

Ron M. & Weller J.I. 2007. From QTL to QTN identification in livestock-winning by points than knock-out: a review. *Animal Genetics* **38**, 429-439.

Roos A.P.W. de, Schrooten C., Mullaart E., Calus M.P.L., Veerkamp R.F. 2007. Breeding value estimation for fat percentage using dense markers on *Bos taurus* autosome 14. *Journal of Dairy Science* **90**, 4821–4829.

Safari E., Fogarty N.M. & Gilmour A.R. 2005. A review of genetic parameter estimates for wool, growth, meat and reproduction traits in sheep. *Livestock Reproduction Science* **92**, 271-289.

SAS release 9.1, SAS Institute, Cary, NC. Ref Type: Computer Program.

Schinckel A.P., Ferrel J., Einstein M.E., Pearce S.M.& Boyd R.D. 2004. Analysis of pig growth from birth to sixty days of age. *Professional Animal Scientist* **20**, 79-86.

Schaeffer L.R. 2004. Application of random regression models in animal breeding. *Livestock Production Science* **86**, 35-45.

Seaton G., Haley C.S., Knott S.A., Kearsey M. & Visscher P.M. 2002. QTL Express: mapping quantitative trait loci in simple and complex pedigrees. *Bioinformatics* **18**, 339-340.

Seaton G., Hernandez J., Grunchec J.A., White I.M.S., Allen J., de Koning D.J. Wei W., Berry D., Haley C.S. & Knott S.A. 2006. GridQTL: a Grid portal for QTL mapping of compute intensive datasets. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production,* Communication No. 27-07 (CD-ROM).

Self S.G & Liang K.Y. 1987. Asymptotic properties of maximum-likelihood estimators and likelihood ratio tests under nonstandard conditions. *Journal of the American Statistical Association* **82**, 605-610.

Short R.E., MacNeil M.D., Grosz M.D., Gerrard D.E. & Grings E.E. 2002. Pleiotropic effects in Hereford, Limousin, and Piedmontese F2 crossbred calves of genes controlling muscularity including the Piedmontese myostatin allele. *Journal of Animal Science* **80**, 1-11.

Sørensen P., Lund M.S., Guldbrandtsen B., Jensen J. & Sorensen D. 2003. A comparison of bivariate and univariate QTL mapping in livestock populations. *Genetics Selection Evolution* **35**, 605-622.

Stearns T.M., Beever J.E., Southey B.R., Ellis M., McKeith F.K.& Rodriguez-Zas S.L. 2005. Evaluation of approaches to detect quantitative trait loci for growth, carcass, and meat quality on swine chromosomes 2, 6, 13, and 18. II. Multivariate and principal component analyses. *Journal of Animal Science* **83**, 2471-2481.

Stone R.T., Keele J.W., Shackelford S.D., Kappes S.M. & Koohmaraie M. 1999. A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *Journal of Animal Science* **77**, 1379-1384.

Stram D.O. & Lee J.W. 1994. Variance-components testing in the longitudinal mixed effects model. *Biometrics* **50**, 1171-1177.

Thompson R., Brotherstone S. & White I.M.S. 2005. Estimation of quantitative genetic parameters. *Philosophical Transactions of the Royal Society B* **360**, 1469-1477.

Turner M.E., Bradley E.L., Kirk K.A. & Pruitt K.M. 1976. A theory of growth. *Mathematical Biosciences* **29**, 367–373.

Vagenas D., White I.M.S., Stear M.J. & Bishop S.C. 2007. Estimation of heritabilities and correlations between repeated faecal egg count measurements in lambs facing natural nematode parasite challenge, using a random regression model. *Journal of Agricultural Science* **145**, 501-508.

Van Laere A.S., Nguyen M., Braunschweig M., Nezer C., Collette C., Moreau L., Archibald A.L., Haley C.S., Buys N., Tally M., Andersson G., Georges M. & Andersson L. 2003. A regulatory mutation in IGF2 causes a major QTL effect on muscle growth in the pig. *Nature* **425**, 832-836.

Varona L., Gòmez-Raya L., Rauw W.M., Ovilo C., Clop A. & Noguera J.L. 2005. The value of prior information for detection of QTL affecting longitudinal traits: an example using Von Bertalanffy growth function. *Journal of Animal Breeding and Genetics* **122**, 37-48.

Verbyla A.P., Cullis B.R., Kenwood M.G. & Welham S.J. 1999. The analysis of designed experiments and longitudinal data by using smoothing splines. *Journal of the Royal Statistical Society. Series C (Applied Statistics)* **48**, 269-311.

Visscher P.M. 2006. A note on the asymptotic distribution of likelihood ratio tests to test variance components. *Twin Research and Human Genetics* **9**, 490-495.

Walling G.A., Visscher P.M., Wilson A.D., Mcteir B.L., Simm G. & Bishop S.C. 2004. Mapping of quantitative trait loci for growth and carcass traits in commercial sheep populations. *Journal of Animal Science* **82**, 2234-2245.

Wang C.S., Rutledge J.J. & Gianola D. 1993. Marginal inferences about variance components in mixed linear model using Gibbs Sampling. *Genetics Selection Evolution* **25**, 41-62.

Wang C.S., Rutledge J.J. & Gianola D. 1994. Bayesian analysis of mixed linear model via Gibbs sampling with application to litter size in Iberian pigs. *Genetics Selection Evolution* **26**, 91-115.

Wang Z. & Zuidhof M.J. 2004. Estimation of growth parameters using a nonlinear mixed Gompertz model. *Poultry Science* **83**, 847-852.

Weller J.I., Wiggans G.R., Van Raden P.M. & Ron M. 1996. Application of a canonical transformation to detection of quantitative trait loci with the aid of genetic markers in a multitrait experiment. *Theoretical and Applied Genetics* **92**, 998-1002.

White I.M.S., Thompson R. & Brotherstone S. 1999. Genetic and environmental smoothing of lactation curves with cubic splines. *Journal of Dairy Science* **82**, 632-638.

Wiener P., Smith J.A., Lewis A.M., Wooliams J.A. & Williams J.L. 2002. Muscle-related traits in cattle: The role of the myostatin gene in the South Devon breed. *Genetics Selection Evolution* **24**, 221-232.

Wilson T., Wu X.Y., Juengel J.L., Ross I.K., Lumsden J.M., Lord E.A., Dodds K.G., Walling G.A., McEwan J.C., O'Connel A.R., McNatty K.P. & Montgomery G.W. 2001. Highly prolific Booroola sheep have a mutation in the intracellular kinase domain of bone morphogenetic protein IB receptor (ALK-6) that is expressed in both oocytes and granulosa cells. *Biology of Reproduction* **64**, 1225-1235.

Wu R.L. & Hou W. 2006. A hyperspace model to decipher the genetic architecture of developmental processes: allometry meets ontogeny. *Genetics* **172**, 627-637.

Wu R.L., Ma C.X., Hou W., Corva P. & Medrano J.F. 2005. Functional mapping of quantitative trait loci that interact with the *hg* mutation to regulate growth trajectories in mice. *Genetics* **171**, 239-249.

Wu R.L., Ma C. X., Lin M. & Casella G. 2004a. A general framework for analyzing the genetic architecture of developmental characteristics. *Genetics* **166**, 1541-1551.

Wu R.L., Wang Z., Zhao W. & Cheverud J.M. 2004b. A mechanistic model for genetic machinery of ontogenetic growth. *Genetics* **168**, 2383-2394.

Wu W.R., Li W.M., Tang D.Z., Lu H.R. & Worland A.J. 1999. Time-related mapping of quantitative trait loci underlying tiller number in rice. *Genetics* **151**, 297-303.

Van der Werf J.H.J., Goddard M. & Meyer K. 1998. The use of covariance functions and random regression for genetic evaluation of milk production based on test day records. *Journal of Dairy Science* **81**, 3300-3308.